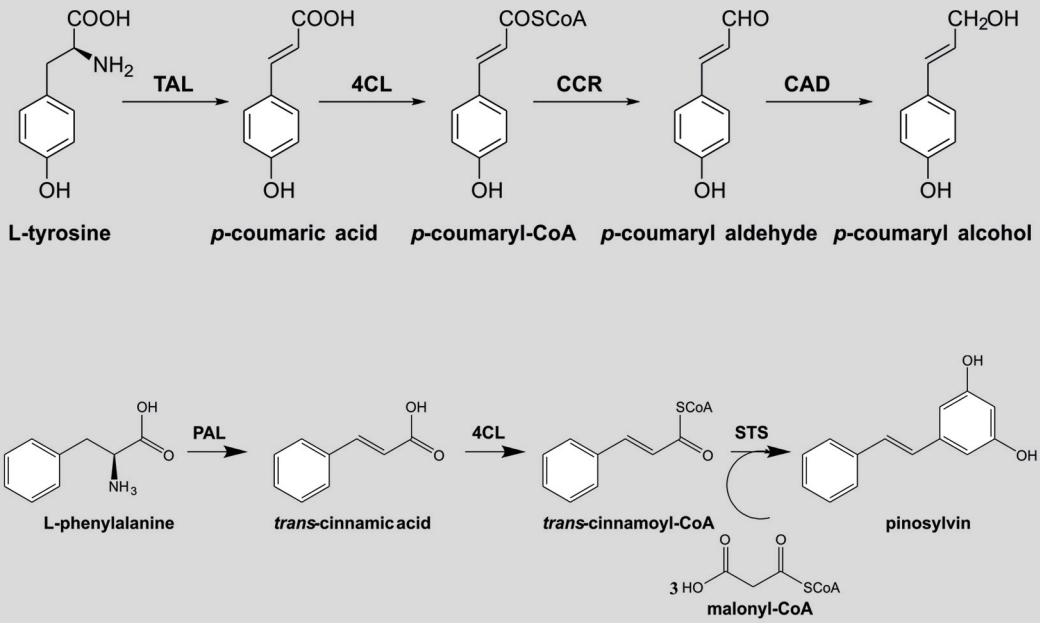
Metabolic engineering of *Escherichia coli* for the production of plant phenylpropanoid derived compounds

Philana Veronica van Summeren-Wesenhagen



Schlüsseltechnologien / Key Technologies
Band / Volume 103
ISBN 978-3-95806-039-5

JÜLICH
FORSCHUNGSZENTRUM

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 103

Forschungszentrum Jülich GmbH
Institute of Bio- and Geosciences
Biotechnology (IBG-1)

Metabolic engineering of *Escherichia coli* for the production of plant phenylpropanoid derived compounds

Philana Veronica van Summeren-Wesenhagen

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies Band / Volume 103

ISSN 1866-1807 ISBN 978-3-95806-039-5

Bibliographic information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Publisher and Distributor: Forschungszentrum Jülich GmbH
Zentralbibliothek
52425 Jülich
Tel: +49 2461 61-5368
Fax: +49 2461 61-6103
Email: zb-publikation@fz-juelich.de
www.fz-juelich.de/zb

Cover Design: Grafische Medien, Forschungszentrum Jülich GmbH

Printer: Grafische Medien, Forschungszentrum Jülich GmbH

Schriften des Forschungszentrums Jülich
Reihe Schlüsseltechnologien / Key Technologies, Band / Volume 103

D 61 (Diss. Düsseldorf, Univ., 2015)

ISSN 1866-1807
ISBN 978-3-95806-039-5

The complete volume is freely available on the Internet on the Jülicher Open Access Server (JuSER) at www.fz-juelich.de/zb/openaccess.

Results presented in this dissertation have been submitted to or published in the following original publications:

van Summeren-Wesenhagen PV and Marienhagen J. (**2013**) Putting bugs to the blush: metabolic engineering for phenylpropanoid-derived products in microorganisms. Bioengineered 4(6):355-362.

van Summeren-Wesenhagen PV and Marienhagen J. (**2014**) Metabolic engineering of *Escherichia coli* for the synthesis of the plant polyphenol pinosylvin. Applied and Environmental Microbiology 81:840-849.

van Summeren-Wesenhagen PV, Voges R, Dennig A, Sokolowsky S, Noack S, Schwaneberg U and Marienhagen J. (**2014**) Combinatorial optimization of synthetic operons for the microbial production of *p*-coumaryl alcohol with *Escherichia coli*. Submitted to ACS Synthetic Biology.

Table of content

Abbreviations

Abbreviations for Système International d'Unités (SI units) of measurement, chemical symbols for the elements and common abbreviations according to international standards, as for example listed in the author guidelines of the "Applied and Environmental Microbiology" journal are not listed in this section.

ACC	acetyl-CoA carboxylase	**GAPDH**	glyceraldehyde-3-phosphate dehydrogenase
At4CL2	4CL2 of *Arabidopsis thaliana N256A/M293P/K320L*	**GBA**	β-glucosidase
AtPAL2	PAL2 of *Arabidopsis thaliana*	**GM**	genetically modified
BRENDA	BRaunschweig ENzyme Database	**2GT**	2-coumarate-O-β-glucosyltransferase
CAD	cinnamyl-alcohol dehydrogenase	**iPCR**	incomplete PCR
CCR	cinnamoyl-CoA reductase	**IDMS-LC-MS/MS**	isotope dilution mass spectrometry coupled to high performance liquid chromatography
C2H	cinnamate 2-hydroxylase	**IPP**	isopentenyl pyrophosphate
C4H	cinnamate 4-hydroxylase	**IPTG**	isopropyl-β-D-thiogalactopyranoside
CFE	cell free extract	**LAC**	laccase
CHI	chalcone isomerase	**LB**	Luria-Bertani
CHS	chalcone synthase	**LC**	liquid chromatograph
4CL	4-coumaryl-CoA ligase	**LIC**	ligation independent cloning
CPEC	circular polymerase extension cloning	**MALDI**	matrix-assisted laser desorption ionization
DIR	dirigent proteins	**MEP**	2C-methyl-D-erythritol-4-phosphate
DMAPP	dimethylallyl pyrophosphate	**MEV**	mevalonic acid
DTT	dithiothreitol	**MON**	cytochrome P450 mono-oxygenases
EDTA	ethylenediaminetetraacetic acid	**MOPS**	3-(N-Morpholino)propanesulfonic acid
epPCR	error prone PCR	**MS**	mass spectrometry
ESI	electrospray ionization	**7-OMA**	7-O-methyl aromadendrin

OMT	O-methyltransferase	**Sc4CL**	4CL of *Streptomyces coelicolor A294G*
PAL	phenylalanine ammonia lyase	**SD**	Shine Dalgarno
PCA	polymerase cycling assembly	**SLICE**	seamless ligation cloning extract
Pc4CL	4CL of *Petroselinum crispum*	**SLIC**	sequence and ligation independent cloning
PcPAL1	PAL1 of *Petroselinum crispum*	**ssDNA**	single stranded DNA
PdSTS3	STS3 of *Pinus densiflora*	**STS**	stilbene synthase
PKS	polyketide synthase	**TAL**	tyrosine ammonia lyase
PLICing	phosphorothioate-based ligase-independent gene cloning	**TCA**	tricarboxylic acid
PNP	plant natural products	**TPS**	terpenoid synthase
PP	phenyl propanoids	**ToF**	time of flight
PPP	pentose phosphate pathway	**uHPLC**	ultra-high-performance LC
PstrSTS2	STS2 of *Pinus strobus*	**USER**	uracil specific excision reagent
PT	prenyl transferases	**YNB**	yeast nitrogen base
RBS	ribosomal binding site	**ZmCAD**	CAD of *Zea mays*
RsTAL	TAL of *Rhodobacter sphaeroides*	**ZmCCR**	CCR of *Zea mays*

Authors' contributions

Putting bugs to the blush: metabolic engineering for phenylpropanoid-derived products in microorganisms.

The manuscript was written by PvS and reviewed by JM.

Metabolic engineering of *Escherichia coli* for the synthesis of the plant polyphenol pinosylvin

PvS and JM designed the study and JM supervised it. The experimental work was performed by PvS. The manuscript was written by PvS and JM.

Combinatorial optimization of synthetic operons for the microbial production of *p*-coumaryl alcohol in *Escherichia coli*

JM designed the project and PvS and AD designed the experiments. PvS performed the main experimental work; RV performed the relative protein determination by IDMS-LC-MS/MS; SS repeated cultivations for constructed strains and SN performed statistical data analysis. All authors contributed to the discussion and interpretation of the obtained results. PvS and JM wrote the manuscript, which was edited by all co-authors.

AD: Alexander Dennig, JM: Jan Marienhagen, PvS: Philana van Summeren-Wesenhagen, RV: Raphael Voges, SN: Stephan Noack, SS: Sascha Sokolowsky.

1. Summary

More than 200,000 plant natural products (PNPs) are currently known. As many of them have favorable effects in the treatment of various diseases such as cancer, diabetes and heart diseases, PNPs are of high interest for the development of new pharmaceutical drugs. A major obstacle during drug development is the limited availability of most PNPs. In comparison to purification from the respective plant sources or costly chemical synthesis, the microbial production of these compounds could provide sufficient quantities from inexpensive substrates.

The main goal of this thesis was the development of a microbial production platform for the synthesis of pinosylvin, a stilbene found in the heartwood of pine trees (e.g. *Pinus strobus* and *Pinus sylvestris*) in *Escherichia coli* from L-phenylalanine. The first step in the biosynthetic pathway is the deamination of L-phenylalanine to *trans*-cinnamic acid, which is catalyzed by a phenylalanine ammonia lyase. The resulting acid is subsequently activated by a 4-coumarate-CoA ligase yielding the coenzyme A thioester *trans*-cinnamoyl-CoA. A stilbene synthase catalyzes the successive condensation of three malonyl-CoA molecules with *trans*-cinnamoyl-CoA and the cyclization of the resulting linear tetraketide intermediate to form the stilbene pinosylvin. Initially, two different variants for each of the three enzymes towards pinosylvin were selected on the basis of their biochemical characteristics. After the construction of several pathway variants and optimization of gene expression, pinosylvin concentrations of up to 3 mg/L were detected by HPLC-MS analysis of ethyl acetate extracts from culture supernatants, showing that the plant pathway was functional in *E. coli*. Analysis of precursor availability and a comparative analysis of intracellular levels of pathway intermediates and product identified limited malonyl-CoA availability and low activity of the pine-derived stilbene synthase as key bottlenecks. By increasing malonyl-CoA availability through addition of the fatty acid synthase inhibitor cerulenin and by *in vivo* evolution of the stilbene synthase from *Pinus strobus* for enhanced activity in *E. coli,* the pinosylvin titer could be increased up to 70 mg/L. A further increase up to 91 mg/L was achieved by supplementation of the precursor L-phenylalanine to the medium.

SUMMARY

The functional integration of synthetic pathways into the microbial host cell metabolism and optimization of heterologous gene expression for achieving high product titers is still a challenging task during metabolic engineering. With the aim to develop a tool for the construction and balancing of synthetic pathways, we advanced the *Phosphorothioate Ligase Independent Gene Cloning* (PLICing) technique and optimized the expression of a polycistronic operon encoding tyrosine ammonia lyase, 4-coumarate-CoA ligase, cinnamoyl-CoA reductase, and cinnamyl alcohol dehydrogenase. These four enzymes catalyze the formation of *p*-coumaryl alcohol from L-tyrosine in *E. coli*. The PLICing method proved to be a fast and efficient technique for the combinatorial assembly of multiple pathway configurations when constructing each pathway configuration individually. A protocol developed for the rapid simultaneous construction of a complete operon library encompassing all possible pathway configurations proved to be unfeasible due to a decreased assembly accuracy of complete operons. Systematic variation of the spacing between the Shine-Dalgarno sequence and the start codon efficiently modulated the translation efficiency of individual pathway genes and helped to balance the synthesis of the *p*-coumaryl alcohol pathway enzymes to achieve high product titers. During the evaluation of 81 pathway variants, tyrosine ammonia lyase activity proved to be the limiting step. Maximization of the translation efficiency of the respective mRNA was decisive for the construction of an *E. coli* strain, which accumulated up to 52 mg/L *p*-coumaryl alcohol in the supernatant.

2. Introduction

2.1. Plant natural products

Plant natural products (PNPs) are plant secondary metabolites that are not essential for plant survival, but aid in the growth and development of the plant in its environment. PNPs play an important role in plant defense, signaling and regulation of metabolic activity within plant cells, but are also responsible for the characteristic colors and fragrances of plants (1).

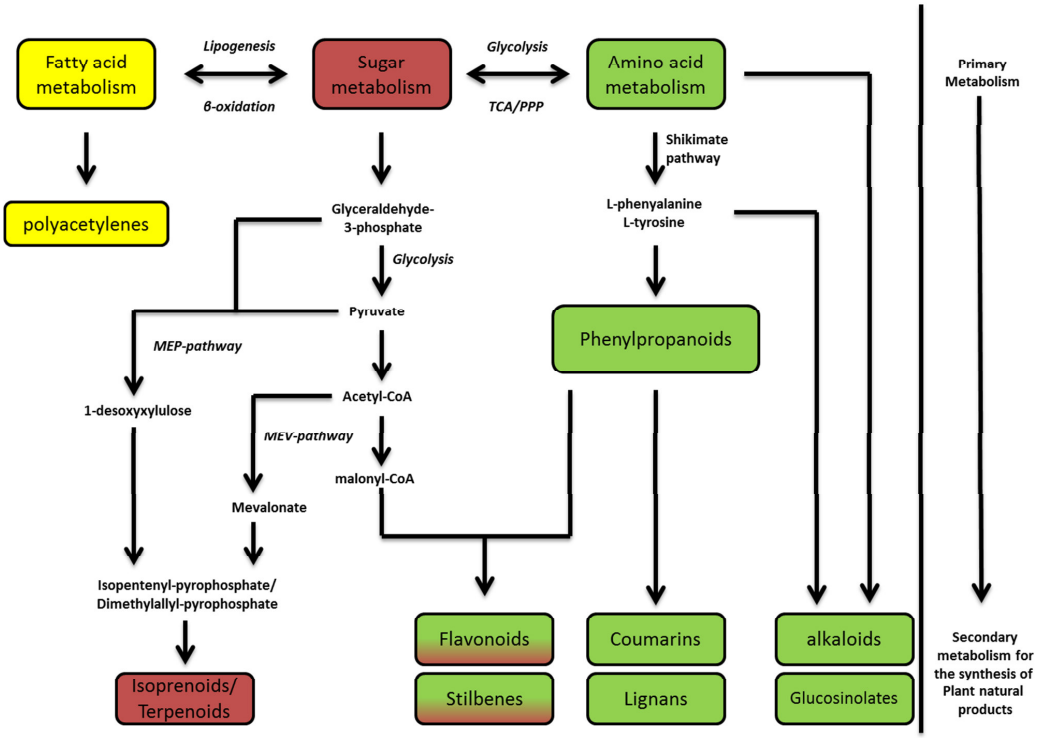


Figure 2.1 Schematic overview of biosynthetic routes leading to the five major classes of plant natural products (PNPs) starting from primary metabolism precursors. Color-code reflects the origin of the different PNP classes from fatty acid-, sugar- or amino acid metabolism.

PNPs are derived from primary metabolites, which are formed during the fatty acid, sugar or amino acid metabolism (Figure 2.1). Basically five major classes of PNPs can be distinguished: isoprenoids, alkaloids, glucosinolates, polyacetylenes and phenylpropanoid and phenylpropanoid-derived polyphenols (1).

2.1.1. Isoprenoids

Isoprenoids also known as terpenoids represent the largest class of PNPs with more than 40,000 compounds estimated to exist in nature (2). Their biological function in nature ranges from light-harvesting in photosynthesis to acting as attractants, repellents, toxins or antibiotics in the context of reproduction, defense or symbiosis with other organisms (3). Plant terpenoids have been widely applied in the pharmaceutical, food, cosmetics, flavors and fragrance, pesticide and disinfectant industry. For example menthol, a monoterpene (C_{10}) that is widely used in the flavor and fragrance industry, has an annual production of 12,000 tons per year (4). Other noteworthy examples of isoprenoids are the antimalarial drug artemisinin, a sesquiterpene (C_{15}) lactone bearing a peroxide bridge isolated from the plant *Artemisia annua* (5) and the anticancer drug taxol, a diterpene (C_{20}) originally isolated from the bark of the Pacific yew, *Taxus brevifolia* (6)(Figure 2.2).

Two pathways are known for the biosynthesis of isoprene building blocks: [1] the mevalonic acid (MEV) pathway, which is utilized by mammals, higher plant species (in the cytosol), archaea and some gram-positive bacteria and [2] the 2C-methyl-D-erythritol-4-phosphate (MEP) pathway, which can be found in higher plants (in the plastids), a few gram-positive bacteria and most gram-negative bacteria (7). Isoprenoids are synthesized by consecutive condensations of the five-carbon isoprene building blocks, isopentenyl pyrophosphate (IPP), with its isomer, dimethylallyl pyrophosphate (DMAPP). Terpenoids are classified on the basis of the number of isoprene units present in their structure. The smallest terpenoids, the hemiterpenoids (C_5) can be formed directly from DMAPP by terpenoid synthase (TPS) activity (8), while consecutive condensation steps of two, three and four isoprene units are catalyzed by prenyl transferases (PT) yielding geranyldiphosphate (GDP; C_{10}), farnesyl diphsophate (FDP; C_{15}) and geranylgeranyl diphosphate (GGDP; C_{20}), respectively (9). TPS enzymes convert GDP, FDP and GGDP into monoterpenoids (C_{10}), sesquiterpenoids (C_{15}) and diterpenoids (C_{20}), in that order. Furthermore, condensation of two molecules of FDP or GGDP gives rise to the precursors of the triterpenoid (C_{30}) and tetraterpenoid (C_{40}) classes. The basic structures of the hemi (C_5)-, mono (C_{10})-, sesqui (C_{15})- and diterpenes (C_{20}), can be further functionalized by various cytochrome P450- dependent mono-oxygenases (P450), reductases, dehydrogenases or numerous classes of transferases, which catalyze oxidations/hydroxylations, reductions, oxidations/dehydrogenations and transfer functional groups (e.g. acyl, aroyl) to the basic structures of terpenoids, respectively (2).

2.1.2. Alkaloids

Alkaloids are low molecular weight nitrogen containing compounds, which are usually derived from amino acids (Figure 2.2). They can be divided into five groups dependent on the amino acid of origin (amino acid or amine pre-cursor is given in brackets):

I. tropane-, pyrrolidine- and pyrrolizide-alkaloids (ornithine),

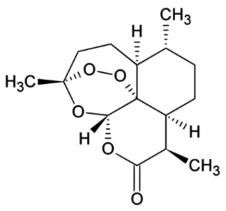
II. benzylisoquinoline (tyrosine),

III. indolequinoline (tryptophan),

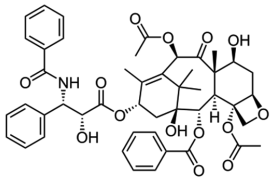
IV. pyridine (pyridine), and

V. quinolizidine- and piperidine-alkaloids (lysine).

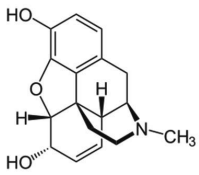
The biosynthetic pathways for alkaloid production usually comprise many enzymatic steps, e.g. 30 enzymes are involved in the biosynthesis of the monoterpenoid indole alkaloid vincristine from tryptophan and geraniol (10). In plants over 12,000 alkaloids have been described to date and most of the known functions of alkaloids are related to plant defense. They often have pharmacological effects and are used as medication (e.g. morphine as analgesic and vincristine as anticancer agent), as recreational drug (cocaine) or stimulant (caffeine and nicotine) (11).



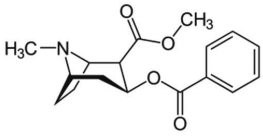
Artemisinin



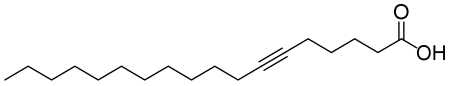
Taxol



Morphine



Cocaine



Tariric acid

Figure 2.2 Exemplary structures of PNPs belonging to the group of isoprenoids (artemisinin and taxol), alkaloids (morphine and cocaine) and polyacetylenes (tariric acid).

2.1.3. Glucosinolates

Glucosinolates represent only a small group of around 120 described structures that contain sulfur and nitrogen and are derived from glucose and an amino acid. Glucosinolates are plant secondary metabolites well-known for their role in plant defense against insects and pathogens, while for humans these compounds function as biopesticides, flavor compounds and also have a large potential as cancer-preventing agents (12).

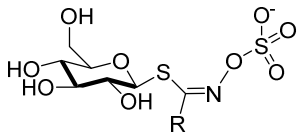


Figure 2.3 Glucosinolate core structure. Variable R group is determined by the side chain of the precursor amino acid.

Upon plant damage glucosinolates are hydrolized to a variety of products that are responsible for all biological activities of this class of compounds. The first hydrolysis step is catalyzed by a thioglucoside glucohydrolase termed myrosinase, leading to the formation of glucose and an unstable aglycone. The aglycone rearranges to form different products, including isothiocyanates, oxazolidine-2-thiones, nitriles, epithionitriles, and thiocyanates, depending on the structure of the amino acid side chain and the presence of additional proteins and cofactors. Glucosinolates can be classified into three groups by their amino acid precursor: [1] compounds derived from L-alanine, L-leucine, L-isoleucine, L-methionine or L-valine are called aliphatic glucosinolates, [2] those derived from L-phenylalanine and L-tyrosine are called aromatic or benzenic glucosinolates, and [3] those derived from L-tryptophan are called indole glucosinolates (12). Glucosinolates have been exclusively found in the order *Brassicales* (*Capparales*) and the family *Putranjivaceae* (*Euphorbiaceae*). The biosynthesis of glucosinolates proceeds through three separate phases: [1] chain elongation of the amino acid side group by methylene ($–CH_2–$) moieties can occur for L-methionine and L-phenylalanine amino acid precursors, [2] formation of the core glucosinolate structure (Figure 2.3, a β-D-glucopyranose residue linked via a sulfur atom to a (Z)-N-hydroximinosulfate ester, plus a variable R group determined by the amino acid side chain) and [3] secondary modification of the amino acid side chain. For aliphatic glucosinolates, secondary modifications include oxygenations, hydroxylations, alkenylations and

benzoylations, whereas indolic glucosinolates, can undergo hydroxylations and methoxylations (13).

2.1.4. Polyacetylenes

To date more than 2,000 polyacetylenes have been identified, which are characterized by their carbon-carbon triple bond(s) or alkynyl functional group(s) (Figure 2.2). Studies concerning their ecological function hint at survival-based roles were they can act as toxin, anti-feedant, allelochemical, phytoalexin or antibiotic (14), but especially their antitumor properties have attracted significant research attention (15). Fatty acids produced during the primary metabolism represent the substrates for polyacetylene biosynthesis. The main mechanism for unsaturated carbon-carbon bond formation in plants is the oxidative dehydrogenation (desaturation) mechanism, catalyzed by non-hemediiron desaturases, which consume molecular oxygen and electrons provided by either NADH or NADPH to excise two vicinal C–H bonds on the alkyl chain of a fatty acid derivative, while simultaneously reducing oxygen to water (14).

2.1.5. Phenylpropanoid and phenylpropanoid-derived polyphenols

Phenylpropanoids (PPs) and PP-derived compounds represent the major source of plant phenolic compounds. PP-derived products can provide protection against ultraviolet light, pathogens and herbivores, but can also serve as plant pigments and fragrances (16). In addition to these qualities, their health promoting effects such as antioxidant, antiviral, antibacterial and anticancer activities, makes them very interesting compounds for the pharmaceutical industry. Furthermore, PP and PP-derived products could be used as aromatic chemical building blocks in the fine chemical, materials synthesis, and flavor and fragrance industries (1).

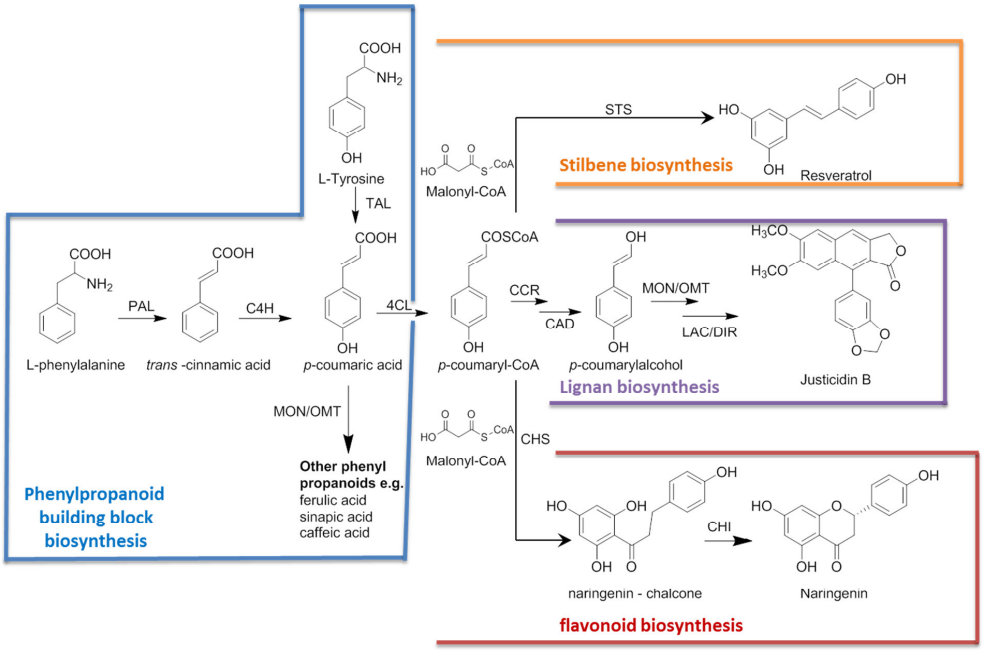


Figure 2.4 Modular overview of biosynthetic pathways for the production of phenylpropanoid precursor structures (blue), stilbenes (orange), lignans (purple) and flavonoids (red). *Abbreviations*: 4CL, 4-coumaryl-CoA ligase; C4H, cinnamate 4-hydroxylase; CAD, cinnamyl alcohol dehydrogenase; CCR, cinnamoyl-CoA reductase; CHI, chalcone isomerase; CHS, chalcone synthase; DIR, dirigent protein; LAC, laccase; MON, cytochrome P450 monooxygenase; OMT, O-methyltransferase; PAL, phenylalanine ammonia lyase; STS, stilbene synthase; TAL, tyrosine ammonia lyase.

Characteristic for the PP structure is their aromatic phenyl group and propene tail, which are derived from the precursor amino acids L-tyrosine and L-phenylalanine. The first step of the phenylpropanoid biosynthesis pathway is catalyzed by a phenylalanine ammonia lyase (PAL) or a tyrosine ammonia lyase (TAL), which converts L-phenylalanine into *trans*-cinnamic acid or L-tyrosine into *p*-coumaric acid through a non-oxidative deamination, respectively (Figure 2.4). *Trans*-cinnamic acid can be converted to *p*-coumaric acid by a cinnamate 4-hydroxylase (C4H). Subsequently, both acids can be modified by Coenzyme A (CoA) activation, reduction, *O*-methylation and/or aromatic hydroxylation to give rise to precursor molecules for the synthesis of flavonoids, stilbenes, coumarins and lignans. Phenylpropanoids are either direct precursors of lignans and coumarins or are incorporated into the stilbene and flavonoid structures, through a hybrid phenylpropanoid/polyketide-pathway requiring malonyl-CoA (Figure 2.4). Whether a stilbene or a flavonoid is formed, depends on the type III polyketide

synthase. Either a stilbene synthase (STS) or a chalcone synthase (CHS), catalyzes the iterative condensation of three malonyl-CoA units to a CoA-activated phenylpropanoic acid (16). However, the resulting tetraketide intermediate is either cyclized via an intramolecular C2 → C7 aldol condensation under direction of the STS to form a stilbene, or an intramolecular C6 →C1 claisen condensation to form a chalcone under guidance of the CHS. An additional ring closing step catalyzed by a chalcone isomerase (CHI), converts the chalcone to the core flavonoid structure. Further modifications such as hydroxylations, methylations, methoxylations, acylations and *C*- and/or *O*-glycosylations give rise to the different classes of flavonoids (16). Quercetin, a flavonol (17) and the stilbene resveratrol (18) have been extensively studied for their antioxidant, anti-inflammatory, anti-cancer and other health promoting effects (Figure 2.5).

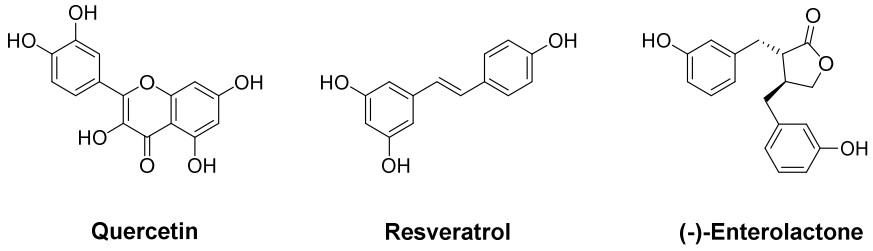


Figure 2.5 The flavonol quercetin, the stilbene resveratrol and (-)-enterolactone, a lignin which is formed from plant lignans by the action of mammalian intestinal bacteria.

A general scheme for coumarin biosynthesis has been inferred, even though the genes encoding the enzymes of even the simplest coumarin structure still have to be identified (Figure 2.6). Generally the coumarin biosynthesis starts from L-phenylalanine via the intermediates *trans*-cinnamate, *trans*-2-coumarate, *trans*-2-coumarate-β-D-glucoside, *cis*-2-coumarate-β-D-glucoside, and *cis*-2-coumarate (19). From a health perspective coumarins have received most attention for their antioxidant properties (20).

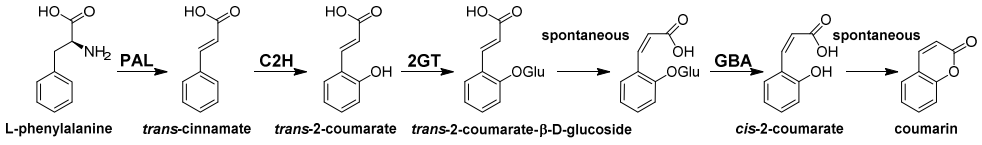


Figure 2.6 Proposed coumarin biosynthetic pathway in plants. *Abbreviations*: PAL, phenylalanine ammonia lyase; C2H, cinnamate 2-hydroxylase; 2GT, 2-coumarate-O-β-glucosyltransferase, GBA, β-glucosidase (19).

Reduction of CoA-activated phenylpropanoic acids leads to the formation of monolignols. Subsequently occurring reductions are catalyzed by a cinnamoyl-CoA reductase (CCR) and a cinnamyl-alcohol dehydrogenase (CAD), while hydroxylases and methyltransferases can modify the intermediates to yield the various monolignol structures (21). Monolignols are precursors of lignin, which give plant cell walls their mechanical strength, and of the structurally diverse group of lignans (22). The main components of lignin are the monolignols coniferyl alcohol, sinapyl alcohol and *p*-coumaryl alcohol (21). Lignans or dilignols are formed through the oxidative dimerization of two monolignol units. Monolignol dimerization is initiated by laccases or peroxidases, resulting in the formation of free radical intermediates. The patterns of dimerization are controlled by so-called dirigent proteins (DIRs). DIRs capture and position the free radicals, to enable formation of only one specific dilignol species (23). Enterolactone (Figure 2.5) formed in the mammalian digestive tract from plant lignans has also been identified as potential anti-cancer agent (24).

2.2. Production of plant natural products

As has been outlined above, PNPs have a multitude of different properties that have made them attractive for the use in pharmaceutical, food, cosmetics, flavors and fragrance, materials synthesis, pesticide and disinfectant industries. For example almost half of all new approved drugs from 1981 to 2010 were (plant) natural products, their semisynthetic derivatives or natural product analogs (25). Sufficient production of PNPs is of great importance for their sustained commercial use. Three strategies can be employed to produce PNPs: (i) extraction from the respective plant sources or from plant cell cultures; (ii) synthetic and semi-synthetic approaches can be employed for the chemical synthesis and (iii) production in genetically engineered microbial host systems.

2.2.1. Plant based production

Solvent extraction of plant material is the most classical method for gaining access to PNPs. Some of the commonly used solvents in the extraction procedures such as hexane, ether, chloroform, acetonitrile, benzene or ethanol are generally used in different ratios with water (26). However, PNPs are usually synthesized in specific types of plant cells (e.g. resveratrol can be found in the skin of grapes (27), but also in the roots of the Japanese knotweed (28)) and only during a certain developmental stage of the plant, or under specific (seasonal)

conditions. The low abundance of most PNPs in natural material and environmental, seasonal, or regional variations of natural sources makes their isolation from plant material difficult and expensive. Downstream processing is further impeded by the presence of structurally similar PNPs. Traditional plant breeding or construction of genetically modified (GM) plants can be employed to increase the abundance of a specific PNP (29). However, commercialization of PNPs through the use of GM plants is often avoided, due to limited customer acceptance and high regulatory hurdles, which must be overcome (30). As an alternative to field crops, plant tissue culture systems such as hairy root cultures and cell suspension cultures can be used, which produce the PNP of interest either constitutively or after elicitation (e.g. UV light, methyljasmonate (31) a signaling molecule derived from plant fatty acids) (32). In comparison to field crops, the genetic modification of tissue cultures is not as strictly regulated by law and cultivation conditions can be more easily controlled (30). Although there are some successful commercial applications of cell suspension cultures (e.g. taxol production by plant cell cultures of *T. chinensis* (33)), the technique still suffers from low and variable product yields, slow growth of cultures, biochemical or genetic instability over time and scale-up issues (33, 34).

2.2.2. Chemical synthesis

Historically, chemical synthesis or semi-synthesis from isolated natural precursors has been used for the production of PNP synthetic analogues, owing to their lack of availability or expensive isolation from natural sources (35). However, increasing customer demand for naturally sourced flavors and colorants in the food industry has diminished the popularity of their synthetic alternatives (36). Furthermore, chemical synthesis mostly relies on petroleum-derived solvents and chemicals, which is a limited resource. The structural complexity of most PNPs, which can include multiple stereo-chemical centers and regio-specific modifications, represents a tremendous challenge for their chemical synthesis even with the advancements in (catalytic) synthetic methods. The complex chemistry of PNPs often requires numerous intermediary protection and re-functionalization steps that do not lead directly to the targeted molecular structure, but ultimately create an amplitude of waste and by-products while the overall yield decreases (37).

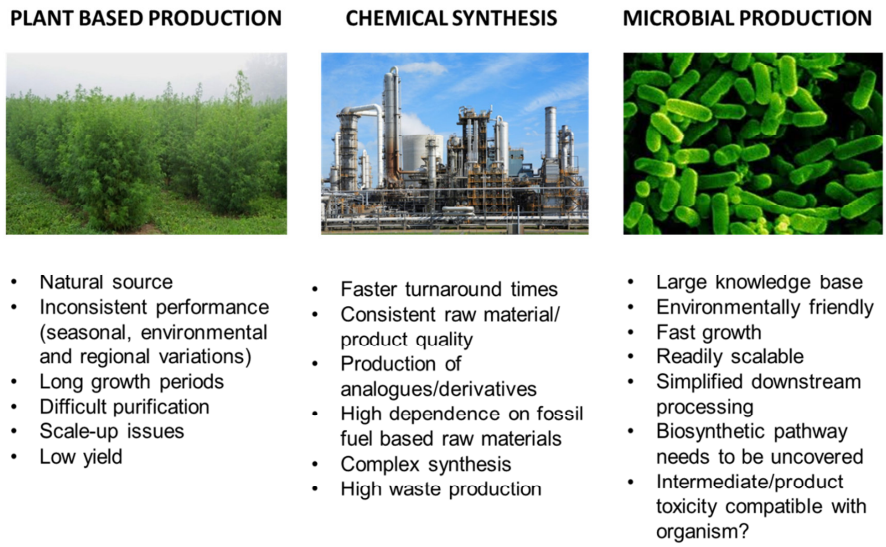


Figure 2.7 Comparison of microbial PNP production to the more traditional plant based PNP production and chemical synthesis of PNPs.

2.2.3. Microbial production

Since the first half of the last century microorganisms have been increasingly used for the production of value-added compounds for the pharmaceutical, chemical, agricultural and food industries (38). The gram negative bacteria *Escherichia coli* and the yeast *Saccharomyces cerevisiae* are two of the most widely used industrial organisms due to their extensively characterized physiology and genetics, high growth rates, and the availability of ample genetic tools. Other well-known industrial platform organisms include *Corynebacterium glutamicum, Bacillus subtilis* and *Pseudomonas putida* (39). The genetic accessibility of these microorganisms allows for the construction of tailor-made recombinant strains by metabolic engineering, which are able to produce compounds of interest. Heterologous production of PNPs is only possible when the metabolic pathway towards the PNP is known, so that the specific pathway genes can be transferred to the heterologous host. Progress in the field of sequencing and recombinant DNA technology has resulted in the elucidation of numerous biosynthetic pathways responsible for the production of PNPs (40). Microbial production of PNPs represents a promising alternative with several advantages in comparison to above-mentioned approaches (Figure 2.7). First, the PNP production proceeds in a benign aqueous environment based on inexpensive renewable

feed stocks, avoiding the use of heavy metals, organic solvents, and strong acids and bases, which are routinely employed during chemical synthesis (1). Additionally, consumer value is increased as flavor and coloring agents produced by microbes are labelled as 'natural' (35). Furthermore, in contrast to natural product synthesis in plants and plant cell cultures, microbial production allows for short production times due their high growth rates. Microbial fermentations are also readily scalable from the lab bench to industrial-sized fermenters. Finally, downstream purification of PNPs is minimized in comparison to plant and plant cell cultures, because in most cases the bacterial host does not have pathways competing with the transgenic plant pathway, resulting in the production of chemically distinct compounds (41).

2.3. Metabolic engineering strategies for microbial PNP strain development

Traditionally, strain development is achieved through iterative rounds of random mutagenesis and selection on the basis of desired phenotypic traits. The advantage of this approach is that strain improvement is possible without knowledge of the biochemical pathways or genetics of the organism, even though this method can be time and resource-intensive and is only applicable for homologous product pathways when no genetic knowledge of the desired biosynthetic pathway is available. With the advent of recombinant DNA technologies and the many advances in the field of DNA sequencing and bioinformatics, it has become possible to transfer complete biosynthetic pathways from native hosts into heterologous microbial organisms. This has paved the way towards microbial PNP production. When choosing a heterologous host for the introduction of a PNP pathway, several points must be taken into account: [1] genetic accessibility of the expression host, [2] availability of biosynthetic precursors, [3] presence of competing metabolic pathways, [4] recognition of codon usage and other genetic elements by the transcription and translation machinery, [5] necessity of post-translational modifications, [6] tolerance against intermediate(s) and product toxicity (42, 43). Multiple metabolic engineering strategies for fine tuning of the resultant recombinant organisms exist, in order to obtain high titers of desired products. These strategies include, increasing precursor and co-factor supply, overexpression of rate limiting enzymes, optimizing and balancing of gene expression, plasmid or genome based expression, removal or reduction of flux to unwanted byproducts or competing pathways, improving product export or a combination of the above (Figure 2.8) (40). With the advent of organism specific metabolic models, scientists have started to

predict genetic modifications, which should maximize the flux toward a desired product. The models take into account all biotransformations present in the production strain and can be used to identify targets for genetic modification that are not directly linked to the biosynthetic pathway of interest (44, 45).

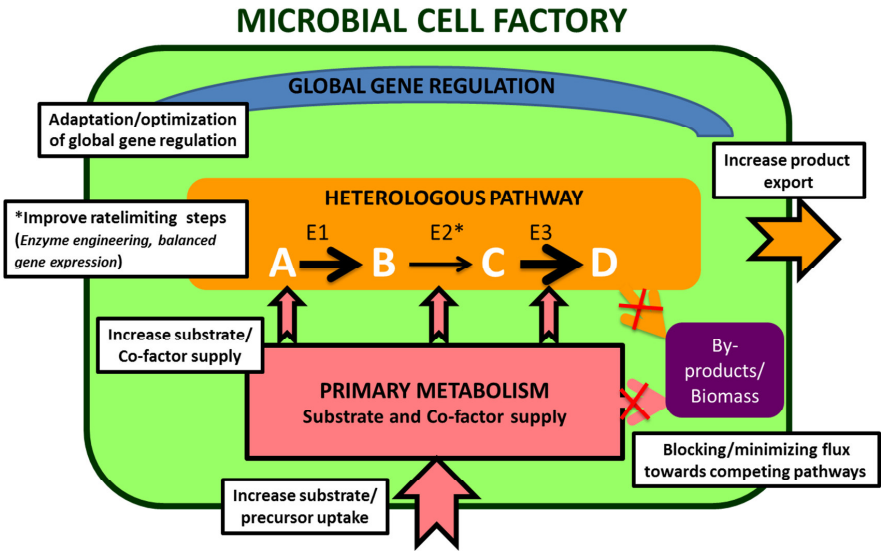


Figure 2.8 Metabolic engineering strategies for improving product titers during microbial biosynthesis. **A** represents the precursor molecule, **B** and **C** the intermediates of the heterologous pathway and **D** corresponds to the product. E1, E2 and E3 represent the heterologously expressed enzymes.

A successful metabolic engineering example for the microbial production of PNPs is the production of the terpenoid artemisinic acid, a precursor of the antimalarial drug artemisinin with *Saccharomyces cerevisiae* (Figure 2.9) (46). Engineering strategies used, included downregulation of the competing pathway for the precursor molecule farnesyl diphosphate, low expression of the reductase (CPR1) and cytochrome b_5 (CYB5) from *Artemisia annua* to improve the activity of the amorphadiene oxidase CYP71AV1 (a cytochrome P450), overexpression of a cytosolic catalase to reduce oxidative stress and changing from a galactose inducible system to constitutive expression of genes involved in the biosynthesis of artemisinic acid. Finally process improvements involving an extractive fermentation with isopropyl myristate oil and pulse-feeding of ethanol resulted in 25 g/L artemisinic acid (46).

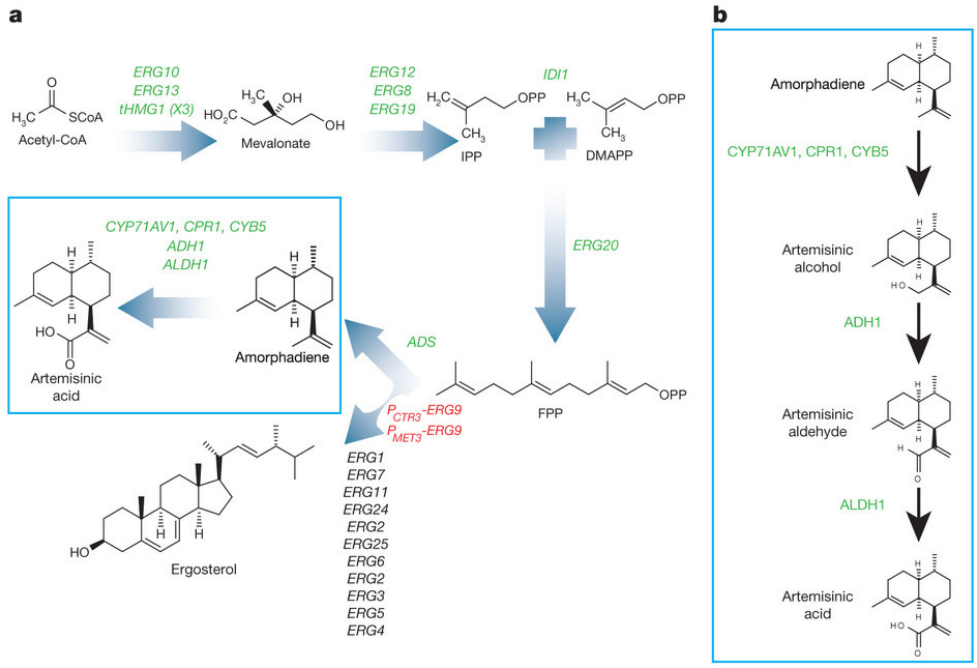


Figure 2.9 (a) General scheme for artemisinic acid production in *Saccharomyces cerevsiae*. Genes changed from a galactose inducible system to a constitutive expression system are shown in green. Copper- or methionine-repressed squalene synthase (ERG9), to downregulate the competing pathway for the precursor molecule farnesyl diphosphate, is shown in red. DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; IPP, isopentenyl diphosphate. tHMG1 encodes truncated HMG-CoA reductase. **(b)** The three-step oxidation pathway of amorphadiene to artemisinic acid from *Artemisia annua* expressed in *S. cerevisiae*. Taken from C.J. Paddon *et al* (46).

2.4. Tools for synthetic pathway assembly

Plant natural product pathways are multi-gene biosynthetic routes and transfer of such pathways to heterologous hosts is greatly facilitated by efficient cloning/assembly techniques. The most commonly used cloning technique relies on the activity of restriction enzymes to generate compatible ends (‘‘sticky or blunt ends”) and a ligase to fuse the DNA fragments that encode the various pathway genes, regulatory elements and expression vector (47). However, for assembling multiple DNA fragments into a single plasmid the method is limited by the availability of unique restriction sites, varied transformation efficiency due to incomplete enzymatic reactions and poor annealing efficiencies of the short single stranded DNA (ssDNA) overhangs (2-4 nucleotides) (48). To address these challenges, enzyme/ligase-free methods have been developed to enable assembly of multiple genetic

components (open reading frames, fusion tags and control elements such as promoters, terminators and ribosome binding sites) in a single step (49, 50). The restriction enzyme/ligase-free methods are generally based on hybridization of ssDNA overhangs or rely on phage recombinases which take advantage of the presence of specific recombination sequences (Gateway cloning (51)) or homologous regions between the fragments (Seamless Ligation Cloning Extract [SLICE] method (52)). The downside of the recombinase based cloning methods, especially of the Gateway cloning method, is that they heavily rely on kits containing special vectors, enzymes, or hosts. The restriction enzyme/ligase-free methods based on hybridization of ssDNA overhangs differ in the procedures leading to single-stranded complementary overhangs of the DNA. Generation of these ssDNA overhangs can be achieved with exonucleases or enzymes with exonuclease activity e.g. T4 DNA polymerase used in Sequence and Ligation Independent Cloning (SLIC) (53) or Ligation Independent Cloning (LIC) (54, 55), exonuclease III (56), T7 gene 6 exonuclease (57) or a uracil-DNA glycosylase combined with endonuclease VIII applied in the Uracil-Specific Excision Reagent (USER) cloning method (58). ssDNA overhangs can also be obtained through incomplete PCR (iPCR) generating unfinished extension products (59). Another method relying exclusively on PCR is the Circular Polymerase Extension Cloning (CPEC), where vector and inserts share overlapping regions at the ends. After denaturation and annealing in each CPEC PCR cycle, the inserts and vector hybridize because of their compatible ends and use each other as a template during PCR extension (Figure 2.10) (60).

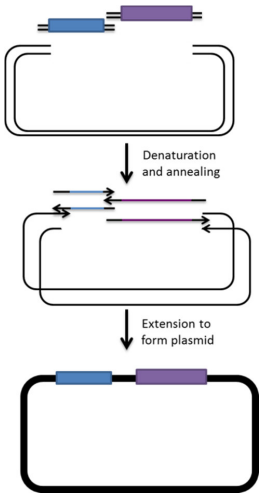


Figure 2.10 Circular Polymerase Extension Cloning (CPEC) principle.

Finally, ssDNA overhangs can be obtained via chemical cleavage when utilizing phosphorothioate modified nucleotides (61, 62). The method also known as Phosphorothioate-based ligase-independent gene cloning (PLICing) starts with the amplification of the insert fragment(s) and the vector by PCR using primers with phosphorothioated nucleotides at the 5'-end. The PCR products are chemically cleaved in an iodine/ethanol buffer, producing single stranded complementary overhangs. The resulting fragments are hybridized at room temperature and can be directly transformed into competent *E. coli* cells, were the natural repair system of this bacterium repairs the single stranded breaks (nicks) in the plasmid at the overlaps between the individual fragments (Figure 2.11) (61, 63).

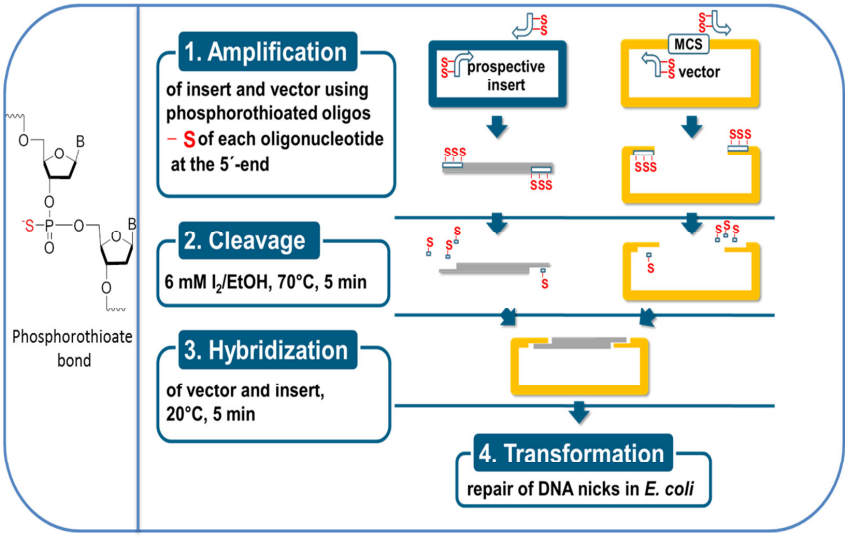


Figure 2.11 DNA backbone with a phosphorothioate bond (left) and schematic overview of the PLICing method composed of four steps: amplification, cleavage, hybridization and transformation (right).

2.5 Aims of this thesis

This dissertation focused on metabolic engineering strategies for the establishment and optimization of phenylpropanoid derived compound production in *E. coli*. Various studies have shown that the stilbene, pinosylvin, could potentially be applied in the treatment of certain cancers, cardiovascular inflammatory diseases and adjuvant arthritis (64-67). The main goal of this work was to establish and optimize the heterologous production of pinosylvin in the platform organism *Escherichia coli*. Multiple genes and

pathway configurations, but also the availability of precursors and intermediates were examined. Finally, laboratory protein evolution was performed to optimize the pinosylvin production from glucose in *E. coli*.

Introduction of heterologous pathways into microbial production strains can be challenging due to an unbalanced expression of the pathway genes, which could potentially lead to the accumulation of toxic intermediates or the formation of inclusion bodies (68). Therefore, an additional aim of this thesis was the development of the *Phosphorothioate ligase independent gene cloning* (PLICing) method to assemble synthetic pathways and balance the heterologous gene expression in *E. coli* (61). To achieve this goal, the PLICing method was used to construct a library of synthetic operons, in which the distance between the Shine Dalgarno (SD) sequence and the translation initiation codon of four genes associated with the biosynthetic production of the monolignol *p*-coumaryl alcohol, was individually varied.

3. Results

The state of the art of metabolic engineering of microorganisms for the synthesis of phenylpropanoid-derived compounds (flavonoids, stilbenes, coumarins and lignans) has been reviewed in the following publication entitled ‘’Putting bugs to the blush: metabolic engineering for phenylpropanoid-derived products in microorganisms’’. The experimental results have been summarized in two publications, one is accepted and one was recently submitted.

In the first paper "Metabolic engineering of *Escherichia coli* for the synthesis of the plant polyphenol pinosylvin", approaches and strategies of metabolic engineering and protein evolution for establishing and optimizing the production of the plant stilbenoid pinosylvin from L-phenylalanine in *E. coli* were employed. At first, two enzymes for each of the three required enzymatic steps from L-phenylalanine to pinosylvin were selected on the basis of published biochemical parameters. The codon optimized genes of the respective enzymes were cloned and the basic expression parameters such as expression temperature, expression time and inducer concentration were optimized. At this point very low product concentrations of 0.068 ± 0.004 mg/L could be measured. Based on these initial experiments, the genetic organization of the synthetic pinosylvin pathway was varied to identify the most suitable pathway configuration. The construct environment was varied through the use of different promoter systems (the T7 promoter from the T7 bacteriophage and the constitutive promoter of the glyceraldehyde-3-phosphate dehydrogenase gene [P_{gap}] from *E. coli* MG1655) and operon or individual expression set-ups. Finally, analysis of precursor, intermediates and, product concentrations during the microbial production of pinosylvin led to the identification of malonyl-CoA and the stilbene synthase (STS) from *Pinus strobus* as bottlenecks during product formation. Shortage of malonyl-CoA could be circumvented by adding cerulenin, which blocks fatty acid synthesis and thus leads to elevated intracellular malonyl-CoA concentrations. Random mutagenesis of the STS gene from *Pinus strobus* and screening for increased product formation yielded an enzyme variant with improved solubility in the heterologous host *E. coli*. Overall, the combination of metabolic engineering strategies and *in vivo* protein evolution in this study enabled the development of an *E. coli* strain, which produced 70 mg/L pinosylvin from glucose.

Additional supplementation of the precursor L-phenylalanine to the medium increased the maximum pinosylvin titer to 91 mg/L.

Other metabolic engineering strategies that were employed, but not mentioned in the publication due to limited applicability for our purpose, are the expression optimization of the stilbene synthase from *Pinus densiflora* (PdSTS3) by using the L-arabinose inducible promoter system (pBAD) and by constructing an N-terminal translational fusion of the STS to the maltose binding protein (MBP) of *E. coli*. By changing the concentration of L-arabinose when using the pBAD system, gene expression levels can be varied and optimized for maximum expression of soluble protein (69), while the MBP protein is known for its ability to enhance the solubility of the tethered heterologous protein (70). However, soluble expression of the PdSTS3 was not improved by either strategy. Furthermore the L-phenylalanine producing strain *E. coli* W3110-4(pF20) in short *E. coli* 4pF20 was used as the host strain for improving malonlyl-CoA production and for the episomal expression of the pinosylvin operon, as both addition of the amino acid precursor L-phenylalanine and the inhibitor of fatty acid biosynthesis cerulenin are still required for improving the microbial synthesis of pinosylvin. *E. coli* 4pF20 is a derivative of *E. coli* K12 with chromosomal deletions Δ(*pheA tyrA aroF*) and harbors plasmid pF20, which is based on pFJF119EH carrying feedback-resistant genes of *pheA* and *aroF* (71). A similar strategy as Zha *et al.* (72) was followed for improving the intracellular malonyl-CoA levels, namely overexpression of the acetyl CoA-carboxylase (ACC) genes *accD1* and *accBC* from *Corynebacterium glutamicum* and the native acetyl-CoA synthase (ACS), while deleting the competing pathways for acetate (*ackA-pta* operon) and ethanol synthesis (*adhE* gene). In order to express target genes cloned in T7 expression vectors in *E. coli* 4pF20, the strain was lysogenized with the λDE3 prophage. While *E. coli* 4pF20 (DE3) removed the need for L-phenylalanine supplementation (similar *trans*-cinnamic acid levels (approx. 200 mg/L) were measured when feeding 3 mM L-phenylalanine to *E. coli* BL21(DE3)), expression of the ACC from *C. glutamicum* abolished the production of pinosylvin. Deletion of the *ackA-pta* operon and replacement of the *adhE gene* with an additional copy of the *acs* gene, also did not lead to more acetyl-CoA production as could be measured from intracellular samples (Figure 3.1, malonyl-CoA levels were below the detection limit). A further strategy that was employed to improve the intracellular malonyl-CoA levels was a non-plasmid based method for up-regulating the activity of the

ACC from *E. coli*. Exchanging the native promoter of the *accBC* operon in the *E. coli* genome with the *LacZYA* operon promoter, results in isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible expression of the biotin carboxyl carrier protein encoded by *accB* and the biotin carboxylase encoded by *accC* (73). However the promoter exchange did not have a positive effect on the pinosylvin titers (malonyl-CoA levels were also below the detection limit).

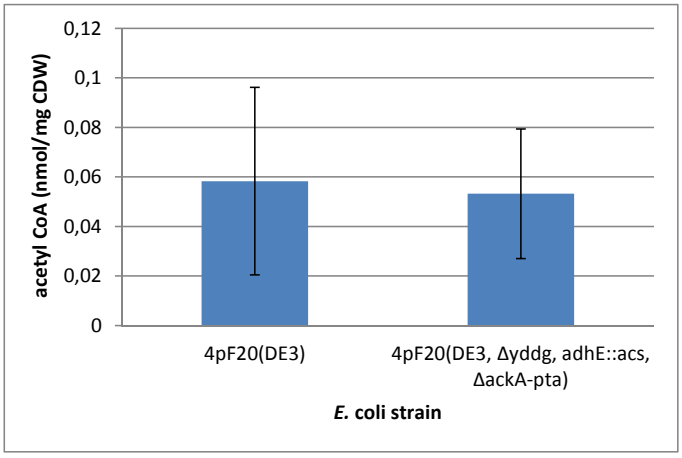


Figure 3.1: Intracellular acetyl-CoA concentration measured for *E. coli* strains 4pF20(DE3) and 4pF20 (DE3, *Δyddg, adhE::acs, ΔackA-pta*).

In the second paper ''Combinatorial optimization of synthetic operons for the microbial production of *p*-coumaryl alcohol in *Escherichia coli*'', the PLICing method was advanced for the rapid combinatorial assembly of multiple pathway configurations in *E. coli*. Subject during the development was a four-step pathway enabling *E. coli* to produce the phenylpropanoid *p*-coumaryl alcohol from the amino acid L-tyrosine. With this method, 81 different pathway configurations could be individually generated, which differed in the length of the spacing between the Shine Dalgarno (SD) sequence and the START codon (5, 9 or 13 nucleotides) in front of each of the four open reading frames in the synthetic operon. An increasing length of this spacing decreased the translation efficiency. Thus, assembly of different gene variants to a four-gene operon allowed for balancing of the heterologous gene expression in *E. coli* and resulted in different product titers of individual strains of the library. The consequences of the modulation of the translation efficiency were reflected by the relative protein amounts in the cells as determined by targeted proteomics applying isotope dilution mass spectrometry coupled to high performance liquid chromatography

(IDMS-LC-MS/MS). In the course of the experiments the tyrosine ammonia lyase from *Rhodobacter sphaeroides* (RsTAL), catalyzing the first step towards *p*-coumaryl alcohol synthesis, was found to be the limiting enzymatic step of the synthetic pathway. For this enzyme the maximum translation efficiency was required for highest *p*-coumaryl alcohol titers (5 nucleotides between the SD sequence and the START codon), whereas medium to low translation efficiency in case of the other three enzymes (9 or 13 nucleotides between the SD sequence and the START codon) was sufficient. The best strain accumulated 52 mg/L *p*-coumaryl alcohol without any further optimization of cultivation conditions.

3.1

COMMENTARY

Bioengineered 4:6, 1–8; November/December 2013; © 2013 Landes Bioscience

Putting bugs to the blush

Metabolic engineering for phenylpropanoid-derived products in microorganisms

Philana V. van Summeren-Wesenhagen and Jan Marienhagen*

Institut für Bio- und Geowissenschaften, IBG-1; Biotechnologie; Forschungszentrum Jülich; Jülich, Germany

Phenylpropanoids and phenylpropanoid-derived phenolic compounds such as flavonoids, anthocyanins or stilbenes are secondary plant metabolites which serve as pigments and scent compounds or provide protection against environmental stress. Due to their antioxidant properties they also have been widely recognized for their benefit on human health. Traditionally, such compounds are extracted from their natural plant sources, but this approach is limited by low abundance and environmental, seasonal as well as regional variations in yield. Chemical synthesis is not a true alternative for the large scale production of more complex phenylpropanoid-derived substances since chemical synthesis becomes commercially unfeasible as the structural complexity of these plant natural products increases. In the last years, many biosynthetic pathways for plant natural products have been elucidated through the advancements in DNA sequencing technologies. In combination with new recombinant DNA technologies this technical progress opens the door toward the functional integration of full biosynthetic pathways for the synthesis of phenylpropanoids and phenylpropanoid-derived compounds in microorganisms. We believe that this approach has great potential to provide sufficient quantities of the desired plant natural product from cheap and renewable resources. This commentary highlights recent advances in the microbial production of phenylpropanoid-derived compounds with an emphasis on flavonoids and stilbenes.

Keywords: phenylpropanoids, flavonoids, anthocyanins, stilbenes, lignans, coumarins, plant natural products, synthetic biology, metabolic engineering, combinatorial biosynthesis

Submitted: 12/20/12

Revised: 02/04/13

Accepted: 02/04/13

http://dx.doi.org/10.4161/bioe.23885

*Correspondence to: Jan Marienhagen; Email: j.marienhagen@fz-juelich.de

Introduction

Phenylpropanoids (PPs) and PP-derived compounds constitute the major source of plant phenolic compounds, which have found numerous applications in food industries as colorants, fragrances and flavoring agents.[1,2] However, today their health promoting qualities are attracting most of the attention. Many studies suggest that various phenylpropanoid-derived compounds possess health-protecting effects against cardiovascular diseases,[3] cancer,[4] diabetes[5] and Alzheimer disease.[6] Not surprisingly, the pharmaceutical industry has already reaped the benefits of the positive health effects of PP-derived compounds and other (plant) natural products: In the area of cancer treatment alone, more than 48% of all novel structures approved as therapeutic agents from around the 1940s to date, were natural products or direct derivatives thereof.[7] Considering the vast chemical diversity of natural products, the percentage of PP-derived compounds and other (plant) natural products among newly discovered drugs is expected to increase in the future.[8] With an increasing demand comes the need for efficient production processes to provide the required quantities for industrial or clinical applications. PPs can be extracted from their native plant sources, but they usually account for less than one percent dry weight of the plant only.[9] This strategy is also limited by slow growth of plants, environmental and regional factors affecting overall yields and difficult separation from structurally similar compounds during purification.

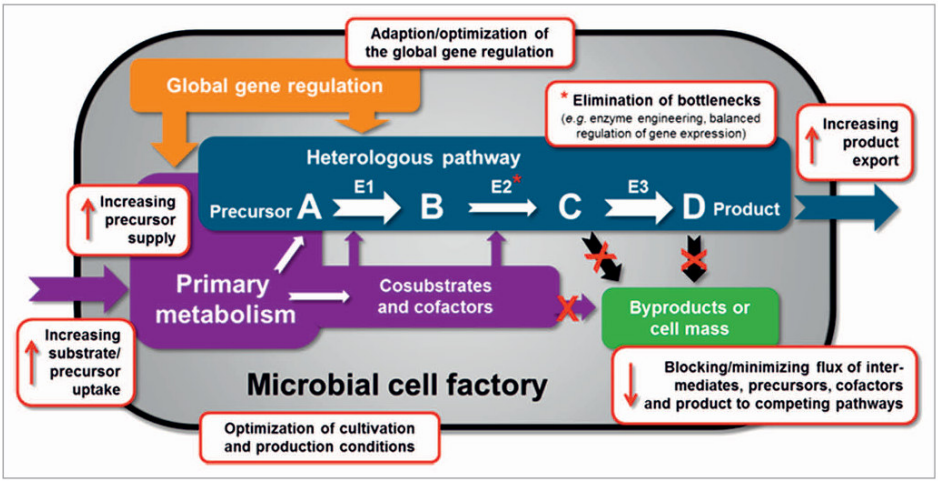


Figure 1. Metabolic engineering strategies for improving product titers. A represents the precursor molecule, B and C the intermediates of the heterologous pathway and D corresponds to the product. E1, E2 and E3 represent the heterologously expressed enzymes.

Although progress has been made in the metabolic engineering of plants for increased synthesis of PPs in the last years, their application as production hosts is still limited.[10,11] Historically, total or semi-synthetic approaches have been used to meet the demand,[12] but the structural complexity of PPs and PP-derived compounds, often possessing multiple chiral centers and functional groups, requires a large number of separate steps during chemical synthesis. Typically, long synthetic routes dramatically decrease the overall yield, while the amount of resources consumed and the number of unwanted products formed increase.[12] In contrast, a recent example shows that plant natural product synthesis with plant cell cultures might prove a promising alternative to isolation and chemical synthesis: Production of Taxol, an anticancer agent originally isolated from the bark of the pacific yew tree (*Taxus brevifolia*), is currently produced in plant cell cultures of *T. chinensis*, reducing the production costs by 80% and 20%, when compared with isolation from *T. chinensis* and semisynthetic production respectively.[13] However, application of plant cell cultures is limited by slow growth of the cell cultures, low and variable product yields, difficult scale-up and complicated downstream processing.[13] In contrast, microbial production of PPs would alleviate many of the aforementioned shortcomings. Well studied microorganisms such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Corynebacterium glutamicum*, *Bacillus subtilis* or *Pseudomonas putida* found already various large-scale applications for the production of numerous compounds in the food, chemical and pharmaceutical industries.[14,15] Microorganisms have the advantage of high growth rates (and thus short production times) and readily scalable cultivation and production technologies. Furthermore, microbial production avoids the use of organic solvents, heavy metals and strong acids or bases making processes more environmentally friendly as compared with chemical production. The main advantage of microorganisms is the availability of extensive molecular tools for their genetic manipulation allowing heterologous expression of whole biosynthetic pathways and the modification of metabolic profiles according to production conditions. In addition, microbial hosts usually provide a clean genetic background for the production of heterologous compounds, as they generally do not possess any competing pathways. As a result, PPs and PP-derived compounds would be synthesized as chemically distinct compounds, which simplifies product purification in comparison to purification from plants or plant cell cultures.[14]

General Metabolic Engineering Strategies for the Microbial Production of Phenylpropanoid-Derived Natural Products

Genetic information of more than 70,000 plants has become available in the last years (http://www.plantgdb.org) and many advances in the field of DNA sequencing and bioinformatics have led to increased knowledge of biosynthetic pathways for secondary metabolites in plants. Identification of genes and enzymes involved in biosynthetic pathways for PP-derived compounds and the development of molecular strategies for the concerted heterologous expression of multiple genes have stimulated rational metabolic engineering of microorganisms for the production of such compounds. In order to improve product yields, many genetic strategies can

be used that optimize the expression of heterologous genes and redirect the metabolic flux toward the production of the desired metabolites (**Fig. 1**). An important decision is the microbial host system, which serves as a chassis for the expression of the heterologous genes. The organism should be genetically accessible to allow the construction of tailor-made recombinant strains and easy to cultivate for production purposes. For the expression of correctly folded enzymes with appropriate kinetic parameters, enzymes can be expressed either from their respective native gene sequence or from a custom-made gene sequence, which has been adapted to the codon usage pattern of the respective microbial host. Protein engineering can be a powerful approach in case individual enzyme parameters turn out to be limiting for the overall flux through the heterologous pathway.[15] Further optimization can be achieved through fine-tuning of individual promoters (constitutive or controlled induction of gene expression) and promoter strength.[16] Alternatively, all or selected genes of the new pathway can be expressed as operon under the control of a single promoter. When using native regulatory elements in the constructed biosynthetic pathway, expression of regulators to stimulate the expression of the pathway genes is also a viable option.[17] In addition, one has to decide whether the heterologous pathway should be expressed plasmid-based or integrated into the host genome. The latter is usually desired in order to develop a stable production strain.[18] The metabolic environment in the host organism can also be optimized for PP production by increasing precursor- and co-factor supply.[19] Furthermore, blocking or minimizing the flux through microbial pathways that also utilize or degrade intermediates or products of the heterologous pathway can result in improved product titers.[19,20] Last but not least, improving product export can be crucial during strain optimization to keep the intracellular product concentrations below cytotoxic levels and to increase the flux through the heterologous pathway.[21] Not surprisingly, almost all of the best performing microbial production strains for the synthesis of PPs and PP-derived compounds were constructed by combining several strategies.

The Phenylpropanoid Pathway

Like all natural products, PPs and derivatives thereof are synthesized from a small subset of primary metabolites provided by the central metabolism. The aromatic phenyl group and the propene tail, which constitute the basic skeleton of all phenylpropanoids are derived from the amino acids L-tyrosine or L-phenylalanine as a result of the non-oxidative elimination of the amino group.[22] This first step of the phenylpropanoid biosynthesis pathway is catalyzed by a phenylalanine ammonia lyase (PAL) or a tyrosine ammonia lyase (TAL), which convert L-phenylalanine into *trans*-cinnamic acid or L-tyrosine into *trans*-coumaric acid, respectively (**Fig. 2**). *Trans*-cinnamic acid can be converted to *trans*-coumaric acid by a cinnamate 4-hydroxylase (C4H). Subsequently, both acids can be modified by reduction, *O*-methylation and/or aromatic hydroxylation to give rise to precursor molecules for flavonoids, stilbenes, coumarins and lignans.

Flavonoids

Flavonoids play important roles in plant pigmentation, fertility and plant defense against UV exposure and pathogen attack.[2] The core structure of the 9000 different flavonoids known to date consists of two aromatic C6 rings interconnected by a three carbon heterocyclic ring. The substitution pattern of this central ring further subdivides flavonoids into flavonols, flavones, flavanones, isoflavones, anthocyanins, flavanonols and flavanols.[2] After activation of (hydroxy)-cinnamic acid to (hydroxy)-cinnamoyl-CoA by 4-coumaryl-CoA ligase (4CL), the first dedicated step toward flavonoid biosynthesis is the consecutive condensation of three malonyl-CoA units with (hydroxy)-cinnamoyl-CoA by a chalcone synthase (CHS) yielding a chalcone (**Fig. 2**). Subsequently, a ring closing step, catalyzed by chalcone isomerase (CHI), yields the core flavonoid structure. Further modifications such as hydroxylations, methylations, methoxylations, acylations and *C*- and/or *O*-glycosylations give rise to the different classes of flavonoids.

Heterologous production of flavonoids in bacteria was first described by Hwang et al. in 2003.[23] The heterologous production of naringenin and pinocembrin in *E. coli* BL21 (DE3) was achieved by placing the three genes encoding a PAL, a 4CL and a CHS individually under control of the inducible T7 promoter. Optimization of culture conditions by supplementing 2 mM L-tyrosine or 2 mM L-phenylalanine led to higher naringenin and pinocembrin titers (452.6 μg/L and 751.9 μg/L respectively).[23] Naringenin production could be improved to a final titer of 20.8 mg/L in TB medium when expressing a TAL from *Rhodobacter sphaeroides* and a 4CL and a CHS from *Arabidopsis thaliana*.[24] Application of a TAL has the advantage that this enzymatic step circumvents the difficult expression of a functional C4H, typically being a membrane bound cytochrome P450 monooxygenase, and its reductase (**Fig. 2**).

Augmenting malonyl-CoA availability by suppression of the fatty acid biosynthesis through the addition of fatty acid synthase inhibitor cerulenin in an *E. coli* strain with enhanced titers of L-tyrosine, resulted in an increase of naringenin production from 29 mg/L to 84 mg/L.[25] In previous studies, flavonoid production was already improved by increasing intracellular malonyl-CoA levels,[21,26,27] confirming the importance of a sufficient supply of this compound during flanovoid synthesis in vivo. A more comprehensive approach to improve intracellular malonyl-CoA levels for increased flavonoid production included the analysis of a genome-scale metabolic network model of *E. coli* and subsequent experimental validation.[28] Bioinformatic analysis pointed toward genetic alterations to upregulate glycolysis in combination with downregulation of the tricarboxylic acid cycle (TCA-cycle) to channel the metabolic flux toward malonyl-CoA. Deletion of the genes for a fumarate hydratase and the β-subunit of the succinyl-CoA synthetase (both TCA-cycle) and overexpression of the genes encoding for the acetyl-CoA carboxylase, phosphoglycerate kinase, glyceraldehyde-3-phosphate

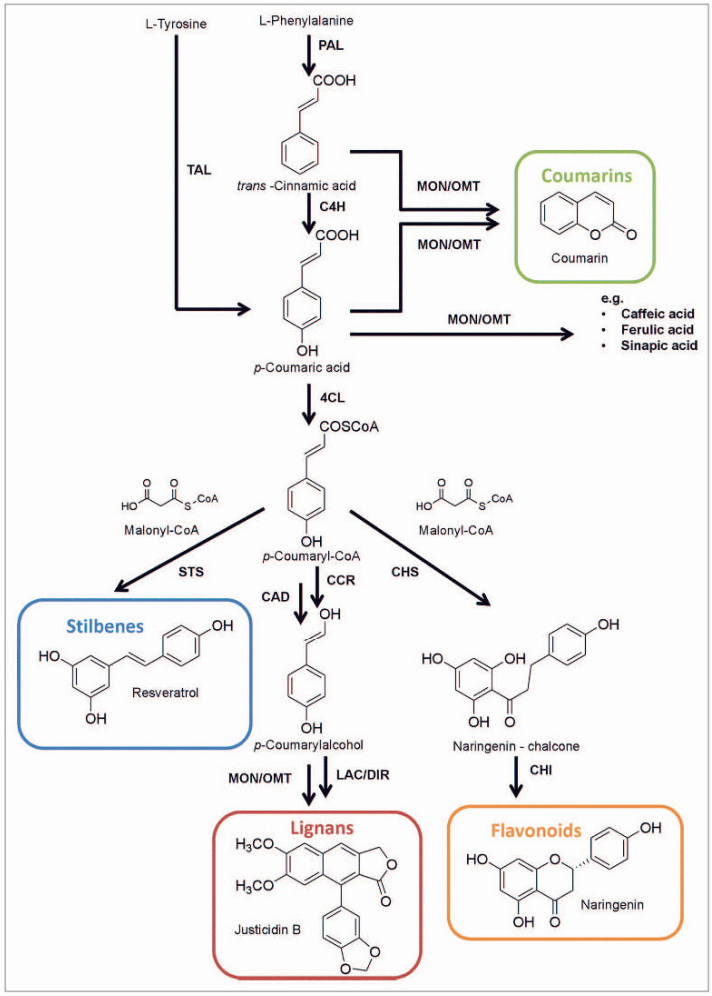


Figure 2. Schematic overview of biosynthetic pathways leading to various phenylpropanoid-derived compounds in plants. Starting either from L-phenylalanine or L-tyrosine, different phenylpropanoids (*p*-coumaric-, caffeic-, ferulic-, sinapic- or *trans*-cinnamic acid) are formed. Phenylpropanoids are precursors of coumarins (green box), stilbenes (blue box), flavonoids (orange box) and lignans (red box). *Abbreviations:* 4CL, 4-coumaryl-CoA ligase; C4H, cinnamate 4-hydroxylase; CAD, cinnamyl alcohol dehydrogenase; CCR, cinnamyl-CoA reductase; CHI, chalcone isomerase; CHS, chalcone synthase; DIR, dirigent protein; LAC, laccase; MON, cytochrome P450 monooxygenase; OMT, *O*-methyltransferase; PAL, phenylalanine ammonia lyase; STS, stilbene synthase; TAL, tyrosine ammonia lyase.

dehydrogenase and the pyruvate dehydrogenase complex resulted in an *E. coli* strain exhibiting a 4-fold increase in intracellular malonyl-CoA levels and accumulating up to 474 mg/L naringenin.[28]

Other important co-substrates/co-factors for the synthesis of flavonoids are UDP-glucose for the synthesis of anthocyanins and NADPH for the biosynthesis of leucoanthocyanidins and (+)-catechins.[29,30]

Increased intracellular UDP-glucose levels could be achieved by overexpressing the nucleoside diphosphate kinase involved in the biosynthesis of UTP from supplemented orotic acid and by

blocking a competing pathway through deletion of the gene for the UDP glucose dehydrogenase rerouting UDP-glucose to UDP-glucuronate.[29] In combination with inhibition of fatty acid synthesis by addition of cerulenin and heterologous expression of a malonate assimilation pathway from *Rhizobium trifolii*, anthocyanin titers of up to 113 mg/L of pelargonidin 3-*O*-glucoside could be achieved with *E. coli*. The importance of other metabolic pathways drawing off precursors and intermediates became also obvious during the heterologous production of the flavonoid 7-*O*-methyl aromadendrin (7-OMA) from p-coumaric acid in *E. coli*. Detailed analysis of the production strains showed that the 7-OMA concentration did not correspond to the precursor consumption, indicating that *p*-coumaric acid is degraded by *E. coli*.[31] Indeed, previous studies showed that *E. coli* is able to utilize aromatic acids as sole carbon source via several degradation pathways.[32] Thus, deletion of these particular metabolic routes represent promising targets for metabolic engineering to improve flavonoid production in *E. coli*.

Stilbenes

Resveratrol, probably the best known stilbene, is believed to prevent cardiovascular diseases and may also provide protection against certain types of cancer, diabetes and neurodegenerative diseases such as Alzheimer disease.[33] These properties make resveratrol an interesting candidate for various applications in the pharmaceutical- and food industry. Similar to the biosynthetic pathway to flavonoids, the pathway to stilbenes branches off from the phenylpropanoid pathway at the stage of the CoA-activated phenylpropanoids (**Fig. 2**). Stilbene synthases (STS), which are also type III polyketide synthases like the CHS, catalyze the formation of the same tetraketide intermediate, but follow a different cyclization mechanism.[34] A STS cyclizes the tetraketide intermediate via an intramolecular C2 → C7 aldol condensation, whereas a CHS follows an intramolecular C6 → C1 Claisen condensation (**Fig. 3**). Microbial production of resveratrol was first reported in *S. cerevisiae*, through the expression of a 4CL from a hybrid poplar (*Populus trichocarpa* × *Populus deltoids*) and a STS from grapevine (*Vitis vinifera*), resulting in 1.45 μg/L of resveratrol starting from *p*-coumaric acid.[35] For optimizing the conversion of *p*-coumaric acid to resveratrol, Zhang et al. employed a fusion protein of 4CL (*A. thaliana*) and STS (*V. vinifera*).[36] Application of this fusion protein resulted in an 8-fold increase of resveratrol compared with a *S. cerevisiae* strain expressing each enzyme individually. It was postulated that the gain in efficiency was either due to channeling of the intermediates from enzyme to enzyme or due to the close proximity of both active sites during catalysis. Various studies reported the production of stilbenes from their phenylpropanoic acid precursors (*p*-coumaric acid and caffeic acid)[37,38] in *E. coli*, employing 4CL and STS from different plant sources (*Nicotiana tabacum* or *A. thaliana* and *Arachis hypagaea* or *V. vinifera*, respectively). Recently a more systematic study was published with respect to the optimization of resveratrol production from *p*-coumaric acid.[39] Two different *E. coli* strains (BL21 Star and BW27784), different promoters (the inducible T7 promoter and the constitutive promoter of the glyceraldehyde-3-phosphate dehydrogenase gene (P_{gapdh})) as well as different gene combinations of a library of two 4CLs and eight STSs were systematically evaluated for improved resveratrol titers. The best performing strain was an *E. coli* BW27784 strain expressing the 4CL from *A. thaliana* and the STS from *V. vinifera* in a bicistronic operon on a pUC plasmid. Production with this strain was improved by inhibiting fatty acid biosynthesis to increase the malonyl-CoA availability to final titers of 2.4 g/L resveratrol. However, final resveratrol titers without precursor feeding and functional expression of a TAL are still quite low (1–2 mg/L).[40,41]

Other Phenylpropanoid-Derived Natural Products

Coumarins, derived from *trans*-cinnamic acid or *p*-coumaric acid (**Fig. 2**) also constitute an important class of active pharmacological substances with possible applications ranging from analgesic to HIV-therapy.[42] Heterologous synthesis of coumarins has not yet been achieved. This can be attributed to insufficient knowledge of the biosynthetic pathways leading to coumarins, which comprise the activity of many enzymes that are difficult to express such as cytochrome P450 monooxygenases or *O*-methyltransferases of unknown identity.

Further reduction of the CoA-activated (hydroxy-)cinnamic acids in the phenylpropanoid pathway finally leads to the formation of their respective alcohols (monolignols) (**Fig. 2**). Monolignols are precursors of lignin, which confers mechanical strength to plant cell walls,[43] and of the structurally diverse group of lignans.[44] Lignans are also known as dilignols because they are formed through the oxidative dimerization of two monolignol units. Dimerization is initiated by laccases or peroxidases forming free radical intermediates, but the patterns of dimerization are controlled by so-called dirigent proteins (DIRs). DIRs capture and orientate the free radicals in such a way as to enable the formation of only one specific dilignol.[45] The guiding role of DIRs is unmistakable as many unspecific coupling products are formed in their absence. Due to the complex glycosylation patterns of the DIRs, their functional expression has been restricted to cell cultures until now.[46,47] Similar to coumarins, knowledge about the biosynthetic pathways leading to complicated lignin structures such as justicin B (**Fig. 2**) is incomplete. In particular, the unknown identity of enzymes involved in further dilignol modifications (cytochrome P450 monooxygenases, dioxygenases, peroxidases and *O*-methyltransferases) have rendered the production of lignans with microorganisms impossible to date.[48]

Heterologous expression of biosynthetic pathways in microorganisms can also be used for the production of unnatural bioactive compounds. This field of research, often referred to as combinatorial biosynthesis, gives access to novel and complex compounds.[49,50] Supplementation of unnatural precursors and/or application of additional enzymes (e.g., hydroxylating, methylating, prenylating enzymes) are the two basic strategies for creating novel bioactive compounds. The relatively

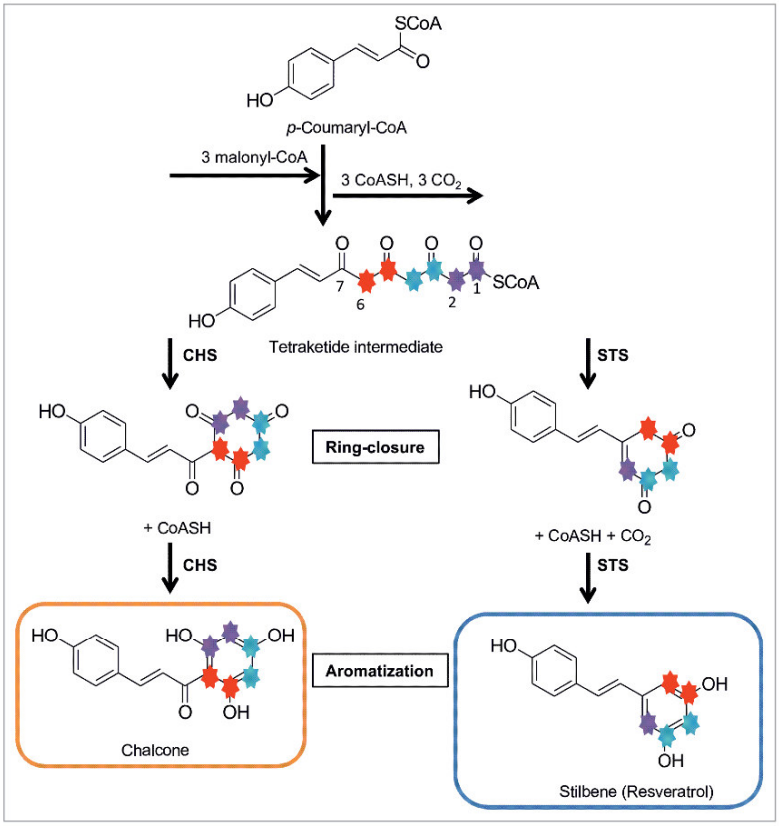


Figure 3. Chalcone synthases (CHS) and stilbene synthases (STS) are type III polyketide synthases, which both catalyze the iterative condensation of three malonyl-CoA units to one CoA-activated phenylpropanoid. However, both enzymes subsequently promote different intramolecular cyclization patterns of the resulting tetraketide intermediate: The STS catalyzes an intramolecular C2 → C7 aldol condensation leading to the typical stilbene scaffold, whereas the CHS catalyzes an intramolecular C6 → C1 Claisen condensation yielding a chalcone.

broad substrate specificity of many plant biosynthetic enzymes enables the conversion of non-natural precursors.[49] For example, Katsuyama and coworkers were able to produce 36 unnatural flavonoids and stilbenes by feeding non-natural carboxylic acids to an *E. coli* strain equipped with a heterologous phenylpropanoid pathway.[50]

Conclusion and Outlook

Metabolic engineering of microorganisms has already proven to be a successful approach to obtain access to phenylpropanoid derived compounds, namely flavonoids and stilbenes. Currently, coumarins and lignans cannot be produced by microorganisms, but considering the dramatic proceedings in DNA sequencing technologies and bioinformatics for the analysis and annotation of (plant) DNA sequences this is expected to change in the future. Additionally, recent developments in the microbial expression of complex eukaryotic enzymes such as plant cytochrome P450 monooxygenases have shown that (rational) protein engineering is a promising strategy for adapting enzymes to the host environment.[14] Furthermore, adaptation of the host metabolism with respect to precursor/co-factor supply has proven its value toward increasing product titers. Further fine-tuning of the host metabolism in relation to the heterologous pathway is expected to enhance product titers even more. As the number of genes in heterologously expressed pathways increases, fast cloning technologies for single genes and synthetic operons will become more important.[51,52] Future technologies should also enable the fine-tuning

of gene expression in order to accomplish an optimal flux through the heterologous pathway for optimized product titers and minimized accumulation of toxic intermediates. In addition to classical strategies such as varying the gene copy number and promoter type, a library of tunable intergenic regions[53] and a design method for Shine-Dalgarno sequences[54] can also be employed to control the relative amounts of individual enzymes on the translational level. Instead of using an existing host in which the heterologous pathway has to be integrated and host and pathway have to be reciprocally adapted, future biotechnologists may directly employ a synthetic producer with a tailor-made metabolic network.[55] Another option is taking advantage of genome-minimized hosts, which only contain genes necessary for growth and heterologous product synthesis.[56] There are still a multitude of levers to pull in the metabolic engineering of microorganisms and undoubtedly many more will be uncovered with the aid of network/metabolic modeling tools, which will unquestionably increase the number of phenylpropanoid-derived plant (non-) natural products that can be produced on a commercially viable scale.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Schwab W, Davidovich-Rikanati R, Lewinsohn E. Biosynthesis of plant-derived flavor compounds. Plant J 2008; 54:712-32; PMID:18476874; http://dx.doi.org/10.1111/j.1365-313X.2008.03446.x.
2. Wang YC, Chen S, Yu O. Metabolic engineering of flavonoids in plants and microorganisms. Appl Microbiol Biotechnol 2011; 91:949-56; PMID:21732240; http://dx.doi.org/10.1007/s00253-011-3449-2.
3. Bradamante S, Barenghi L, Villa A. Cardiovascular protective effects of resveratrol. Cardiovasc Drug Rev 2004; 22:169-88; PMID:15492766; http://dx.doi.org/10.1111/j.1527-3466.2004.tb00139.x.
4. Kale A, Gawande S, Kotwal S. Cancer phytotherapeutics: role for flavonoids at the cellular level. Phytother Res 2008; 22:567-77; PMID:18398903; http://dx.doi.org/10.1002/ptr.2283.
5. Wedick NM, Pan A, Cassidy A, Rimm EB, Sampson L, Rosner B, et al. Dietary flavonoid intakes and risk of type 2 diabetes in US men and women. Am J Clin Nutr 2012; 95:925-33; PMID:22357723; http://dx.doi.org/10.3945/ajcn.111.028894.
6. Pallas M, Canudas AM, Junyent F, Jimenez A, Verdaguer E, Camins A. Resveratrol: A therapeutic approach to neurodegenerative diseases and aging. Mini Rev Org Chem 2010; 7:267-71; http://dx.doi.org/10.2174/157019310792246391.
7. Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 2012; 75:311-35; PMID:22316239; http://dx.doi.org/10.1021/np200906s.
8. Harvey AL, Clark RL, Mackay SP, Johnston BF. Current strategies for drug discovery through natural products. Expert Opin Drug Discov 2010; 5:559-68; PMID:22823167; http://dx.doi.org/10.1517/17460441.2010.488263.
9. Georgiev MI, Weber J, Maciuk A. Bioprocessing of plant cell cultures for mass production of targeted compounds. Appl Microbiol Biotechnol 2009; 83:809-23; PMID:19488748; http://dx.doi.org/10.1007/s00253-009-2049-x.
10. Wu SQ, Chappell J. Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. Curr Opin Biotechnol 2008; 19:145-52; PMID:18375112; http://dx.doi.org/10.1016/j.copbio.2008.02.007.
11. Pollier J, Moses T, Goossens A. Combinatorial biosynthesis in plants: a (p)review on its potential and future exploitation. Nat Prod Rep 2011; 28:1897-916; PMID:21952724; http://dx.doi.org/10.1039/c1np00049g.
12. Newhouse T, Baran PS, Hoffmann RW. The economies of synthesis. Chem Soc Rev 2009; 38:3010-21; PMID:19847337; http://dx.doi.org/10.1039/b821200g.
13. Wilson SA, Roberts SC. Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. Plant Biotechnol J 2012; 10:249-68; PMID:22059985; http://dx.doi.org/10.1111/j.1467-7652.2011.00664.x.
14. Chemler JA, Koffas MAG. Metabolic engineering for plant natural product biosynthesis in microbes. Curr Opin Biotechnol 2008; 19:597-605; PMID:18992815; http://dx.doi.org/10.1016/j.copbio.2008.10.011.
15. Kaul P, Asano Y. Strategies for discovery and improvement of enzyme function: state of the art and opportunities. Microb Biotechnol 2012; 5:18-33; PMID:21883976; http://dx.doi.org/10.1111/j.1751-7915.2011.00280.x.
16. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 2006; 72:211-22; PMID:16791589; http://dx.doi.org/10.1007/s00253-006-0465-8.
17. Chen Y, Wendt-Pienkowski E, Shen B. Identification and utility of FdmR1 as a *Streptomyces* antibiotic regulatory protein activator for fredericamycin production in *Streptomyces griseus* ATCC 49344 and heterologous hosts. J Bacteriol 2008; 190:5587-96; PMID:18556785; http://dx.doi.org/10.1128/JB.00592-08.
18. Chiang CJ, Chen PT, Chao YP. Replicon-free and markerless methods for genomic insertion of DNAs in phage attachment sites and controlled expression of chromosomal genes in *Escherichia coli*. Biotechnol Bioeng 2008; 101:985-95; PMID:18553504; http://dx.doi.org/10.1002/bit.21976.
19. Zha WJ, Rubin-Pitel SB, Shao ZY, Zhao HM. Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. Metab Eng 2009; 11:192-8; PMID:19558964; http://dx.doi.org/10.1016/j.ymben.2009.01.005.
20. Xie XK, Wong WW, Tang Y. Improving simvastatin bioconversion in *Escherichia coli* by deletion of *bioH*. Metab Eng 2007; 9:379-86; PMID:17625941; http://dx.doi.org/10.1016/j.ymben.2007.05.006.
21. Miyahisa I, Kaneko M, Funa N, Kawasaki H, Kojima H, Ohnishi Y, et al. Efficient production of (2S)-flavanones by *Escherichia coli* containing an artificial biosynthetic gene cluster. Appl Microbiol Biotechnol 2005; 68:498-504; PMID:15770480; http://dx.doi.org/10.1007/s00253-005-1916-3.
22. Vogt T. Phenylpropanoid biosynthesis. Mol Plant 2010; 3:2-20; PMID:20035037; http://dx.doi.org/10.1093/mp/ssp106.
23. Hwang EI, Kaneko M, Ohnishi Y, Horinouchi S. Production of plant-specific flavanones by *Escherichia coli* containing an artificial gene cluster. Appl Environ Microbiol 2003; 69:2699-706; PMID:12732539; http://dx.doi.org/10.1128/AEM.69.5.2699-2706.2003.
24. Watts KT, Lee PC, Schmidt-Dannert C. Exploring recombinant flavonoid biosynthesis in metabolically engineered *Escherichia coli*. Chembiochem 2004; 5:500-7; PMID:15185374; http://dx.doi.org/10.1002/cbic.200300783.
25. Santos CNS, Koffas M, Stephanopoulos G. Optimization of a heterologous pathway for the production of flavonoids from glucose. Metab Eng 2011; 13:392-400; PMID:21320631; http://dx.doi.org/10.1016/j.ymben.2011.02.002.
26. Fowler ZL, Gikandi WW, Koffas MAG. Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. Appl Environ Microbiol 2009; 75:5831-9; PMID:19633125; http://dx.doi.org/10.1128/AEM.00270-09.
27. Leonard E, Lim KH, Saw PN, Koffas MAG. Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. Appl Environ Microbiol 2007; 73:3877-86; PMID:17468269; http://dx.doi.org/10.1128/AEM.00200-07.
28. Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MAG. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab Eng 2011; 13:578-87; PMID:21763447; http://dx.doi.org/10.1016/j.ymben.2011.06.008.
29. Leonard E, Yan Y, Fowler ZL, Li Z, Lim CG, Lim KH, et al. Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. Mol Pharm 2008; 5:257-65; PMID:18333619; http://dx.doi.org/10.1021/mp7001472.
30. Chemler JA, Fowler ZL, McHugh KP, Koffas MAG. Improving NADPH availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering. Metab Eng 2010; 12:96-104; PMID:19628048; http://dx.doi.org/10.1016/j.ymben.2009.07.003.
31. Malla S, Koffas MAG, Kazlauskas RJ, Kim BG. Production of 7-O-methyl aromadendrin, a medicinally valuable flavonoid, in *Escherichia coli*. Appl Environ Microbiol 2012; 78:684-94; PMID:22101053; http://dx.doi.org/10.1128/AEM.06274-11.
32. Díaz E, Ferrández A, Prieto MA, García JL. Biodegradation of aromatic compounds by *Escherichia coli*. Microbiol Mol Biol Rev 2001; 65:523-69; PMID:11729263; http://dx.doi.org/10.1128/MMBR.65.4.523-569.2001.
33. Kiselev KV. Perspectives for production and application of resveratrol. Appl Microbiol Biotechnol 2011; 90:417-25; PMID:21360145; http://dx.doi.org/10.1007/s00253-011-3184-8.
34. Austin MB, Bowman ME, Ferrer JL, Schröder J, Noel JP. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. Chem Biol 2004; 11:1179-94; PMID:15380179; http://dx.doi.org/10.1016/j.chembiol.2004.05.024.
35. Becker JVW, Armstrong GO, van der Merwe MJ, Lambrechts MG, Vivier MA, Pretorius IS. Metabolic engineering of *Saccharomyces cerevisiae* for the synthesis of the wine-related antioxidant resveratrol. FEMS Yeast Res 2003; 4:79-85; PMID:14554199; http://dx.doi.org/10.1016/S1567-1356(03)00157-0.
36. Zhang YS, Li SZ, Li J, Pan XQ, Cahoon RE, Jaworski JG, et al. Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and Mammalian cells. J Am Chem Soc 2006; 128:13030-1; PMID:17017764; http://dx.doi.org/10.1021/ja0622094.

37. Beekwilder J, Wolswinkel R, Jonker H, Hall R, de Vos CHR, Bovy A. Production of resveratrol in recombinant microorganisms. Appl Environ Microbiol 2006; 72:5670-2; PMID:16885328; http://dx.doi.org/10.1128/AEM.00609-06.
38. Watts KT, Lee PC, Schmidt-Dannert C. Biosynthesis of plant-specific stilbene polyketides in metabolically engineered *Escherichia coli*. BMC Biotechnol 2006; 6:22; PMID:16551366; http://dx.doi.org/10.1186/1472-6750-6-22.
39. Lim CG, Fowler ZL, Hueller T, Schaffer S, Koffas MAG. High-yield resveratrol production in engineered *Escherichia coli*. Appl Environ Microbiol 2011; 77:3451-60; PMID:21441338; http://dx.doi.org/10.1128/AEM.02186-10.
40. Wang YC, Halls C, Zhang J, Matsuno M, Zhang YS, Yu O. Stepwise increase of resveratrol biosynthesis in yeast *Saccharomyces cerevisiae* by metabolic engineering. Metab Eng 2011; 13:455-63; PMID:21570474; http://dx.doi.org/10.1016/j.ymben.2011.04.005.
41. Choi O, Wu CZ, Kang SY, Ahn JS, Uhm TB, Hong YS. Biosynthesis of plant-specific phenylpropanoids by construction of an artificial biosynthetic pathway in *Escherichia coli*. J Ind Microbiol Biotechnol 2011; 38:1657-65; PMID:21424580; http://dx.doi.org/10.1007/s10295-011-0954-3.
42. Wu L, Wang X, Xu W, Farzaneh F, Xu R. The structure and pharmacological functions of coumarins and their derivatives. Curr Med Chem 2009; 16:4236-60; PMID:19754420; http://dx.doi.org/10.2174/092986709789578187.
43. Neutelings G. Lignin variability in plant cell walls: contribution of new models. Plant Sci 2011; 181:379-86; PMID:21889043; http://dx.doi.org/10.1016/j.plantsci.2011.06.012.
44. Korkina L, Kostyuk V, De Luca C, Pastore S. Plant phenylpropanoids as emerging anti-inflammatory agents. Mini Rev Med Chem 2011; 11:823-35; PMID:21762105; http://dx.doi.org/10.2174/138955711796575489.
45. Davin LB, Lewis NG. Dirigent phenoxy radical coupling: advances and challenges. Curr Opin Biotechnol 2005; 16:398-406; PMID:16023845; http://dx.doi.org/10.1016/j.copbio.2005.06.010.
46. Kim MK, Jeon JH, Fujita M, Davin LB, Lewis NG. The western red cedar (Thuja plicata) 8-8' DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. Plant Mol Biol 2002; 49:199-214; PMID:11999375; http://dx.doi.org/10.1023/A:1014940930703.
47. Pickel B, Constantin MA, Pfannstiel J, Conrad J, Beifuss U, Schaller A. An enantiocomplementary dirigent protein for the enantioselective laccase-catalyzed oxidative coupling of phenols. Angew Chem Int Ed Engl 2010; 49:202-4; PMID:19946920; http://dx.doi.org/10.1002/anie.200904622.
48. Hemmati S, von Heimendahl CBI, Klaes M, Alfermann AW, Schmidt TJ, Fuss E. Pinoresinol-lariciresinol reductases with opposite enantiospecificity determine the enantiomeric composition of lignans in the different organs of *Linum usitatissimum L*. Planta Med 2010; 76:928-34; PMID:20514607; http://dx.doi.org/10.1055/s-0030-1250036.
49. Schwab W. Metabolome diversity: too few genes, too many metabolites? Phytochemistry 2003; 62:837-49; PMID:12590111; http://dx.doi.org/10.1016/S0031-9422(02)00723-9.
50. Katsuyama Y, Funa N, Miyahisa I, Horinouchi S. Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in *Escherichia coli*. Chem Biol 2007; 14:613-21; PMID:17584609; http://dx.doi.org/10.1016/j.chembiol.2007.05.004.
51. Blanusa M, Schenk A, Sadeghi H, Marienhagen J, Schwaneberg U. Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. Anal Biochem 2010; 406:141-6; PMID:20646988; http://dx.doi.org/10.1016/j.ab.2010.07.011.
52. Shao ZY, Zhao H, Zhao HM. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 2009; 37:e16; PMID:19074487; http://dx.doi.org/10.1093/nar/gkn991.
53. Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol 2006; 24:1027-32; PMID:16845378; http://dx.doi.org/10.1038/nbt1226.
54. Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 2009; 27:946-50; PMID:19801975; http://dx.doi.org/10.1038/nbt.1568.
55. Lartigue C, Vashee S, Algire MA, Chuang RY, Benders GA, Ma L, et al. Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science 2009; 325:1693-6; PMID:19696314; http://dx.doi.org/10.1126/science.1173759.
56. Pósfai G, Plunkett G 3rd, Fehér T, Frisch D, Keil GM, Umenhoffer K, et al. Emergent properties of reduced-genome *Escherichia coli*. Science 2006; 312:1044-6; PMID:16645050; http://dx.doi.org/10.1126/science.1126439.

3.2

Metabolic engineering of *Escherichia coli* for the synthesis of the plant polyphenol pinosylvin

Philana van Summeren-Wesenhagen and Jan Marienhagen[#]

Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich D-52425, Germany

Plant polyphenols are of great interest for drug discovery and drug development since many of these compounds have health promoting activities against various diseases such as diabetes, cancer or heart diseases. However, the limited availability of polyphenols represents a major obstacle to be overcome towards clinical applications. In comparison to isolation from natural sources or costly chemical synthesis, the microbial production of these compounds could provide sufficient quantities from inexpensive substrates. In this work, we describe the development of an *Escherichia coli* platform strain for the production of pinosylvin, a stilbene found in the heartwood of pine trees, which could aid in the treatment of various cancers and cardiovascular diseases. Initially, several configurations of the three-step biosynthetic pathway to pinosylvin were constructed from a set of two different enzymes for each enzymatic step. After optimization of gene expression and evaluation of different construct environments, low pinosylvin concentrations up to 3 mg/L could be detected. Analysis of precursor supply and a comparative analysis of the intracellular pools of pathway intermediates and product identified the limited malonyl-CoA availability and low stilbene synthase activity in the heterologous host as main bottlenecks during pinosylvin production. Addition of cerulenin for increasing intracellular malonyl-CoA pools and *in vivo* evolution of the stilbene synthase from *Pinus strobus* for an improved activity in *E. coli* proved to be the keys to elevated product titers. These measures allowed product titers of 70 mg/L pinosylvin from glucose, which could be further increased to 91 mg/L by the addition of L-phenylalanine.

Plant polyphenols are plant secondary metabolites, which have important functions in plant defense and signaling or serve as pigments. Furthermore, many polyphenols have anti-oxidant, anti-cancer and anti-inflammatory properties, rendering these metabolites an invaluable source of bioactive compounds for drug discovery and drug development in the pharmaceutical industry (1).

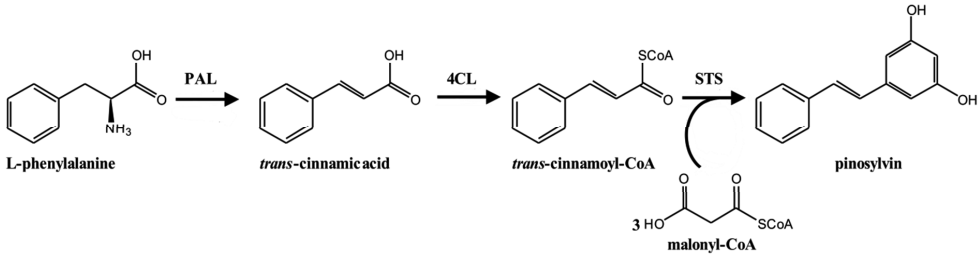


FIG 1 Biosynthetic pathway from L-phenylalanine to pinosylvin. Abbreviations: PAL: phenylalanine ammonia lyase, 4CL: 4-coumarate-CoA ligase, STS: stilbene synthase.

The polyphenol pinosylvin (*trans*-3,5-dihydroxystilbene) is a stilbene, which can be predominantly found in the heartwood of conifer trees of the genus *Pinus* (2). Pinosylvin protects the plants against microbial and fungal decay (3), but has also attracted attention due to its health benefits for humans. Several recent studies indicate a positive effect of pinosylvin in the treatment of various cancers (4, 5), cardiovascular inflammatory diseases (6) and adjuvant arthritis (7). Pinosylvin is synthesized from the aromatic amino acid L-phenylalanine in three enzymatic steps (Fig. 1) (8). L-phenylalanine is first deaminated to *trans*-cinnamic acid by a phenylalanine ammonia lyase (PAL, EC 4.3.1.24). The resulting acid is subsequently coenzyme A (CoA)-activated by a 4-Coumarate-CoA ligase (4CL, EC 6.2.1.12) yielding the CoA thioester *trans*-cinnamoyl-CoA. A stilbene synthase (STS, EC 2.3.1.146), which is a type III polyketide synthase (PKS), catalyzes the successive condensation of three malonyl-CoA molecules with *trans*-cinnamoyl-CoA, forming a linear tetraketide intermediate. Finally, the STS also cyclizes the tetraketide intermediate via an intramolecular C2 → C7 aldol condensation to form the stilbene pinosylvin (8).

The concentration of pinosylvin in pine heartwood ranges from 1 – 40 mg/g dry weight (including its monomethyl ether) (3), rendering the access to pinosylvin through extraction difficult. Furthermore, purification of pinosylvin from these extracts would require the separation from a multitude of often structurally very similar compounds such as other stilbenes or flavonoids (9). Various catalytic and non-catalytic chemical synthetic routes have been devised for the synthesis of stilbenes. However these strategies suffer from the demand for expensive substrates, catalyst instability or degradation or the formation of by-products and various isomers (10). In contrast, microbial production of pinosylvin from glucose would have the advantage of being much more environmentally friendly compared to chemical synthesis, since it avoids the use of organic solvents, heavy metals and strong acids or bases. To date, several microorganisms have been engineered for

the microbial stilbene production, especially for the synthesis of resveratrol (11-13). However, the stilbene production with these strains relies on the supplementation of phenylpropanoids (e.g. coumaric acid or cinnamic acid) as stilbene precursors. This strategy has led to some considerable success with regard to the stilbene resveratrol, where feeding of 15 mM *p*-coumaric acid resulted in maximum product titers of 1.4 g/L in *E. coli* (11), which could be further improved to 1.6 g/L by increasing the carbon flux into malonyl-CoA (14). More recently, *E. coli* has been also engineered to produce 35 mg/L resveratrol from 3 mM of its amino acid precursor L-tyrosine (15). This was achieved by following a modular metabolic

TABLE 1 Strains and plasmids used in this study.

Strain or plasmid	Relevant characteristics	Source or reference
Strains		
E. coli DH5α	F^- ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rk^-, mk^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-	Invitrogen (Karlsruhe, Germany)
E. coli BL21(DE3)	F^- *omp*T *hsd*SB(rB^-, mB^-) *gal dcm* (DE3)	Invitrogen (Karlsruhe, Germany)
Plasmids		
pCDFDuet1	Spt^R; 2x T7*lac* promoters, CDF replicon, *lacI*, His_6-tag, S-tag	Merck Millipore (Billerica, MA, USA)
pETDuet1	Amp^R; 2x T7*lac* promoters, ColE1 replicon, *lacI*, His_6-tag, S-tag	Merck Millipore (Billerica, MA, USA)
pRSFDuet1	Kan^R; 2x T7*lac* promoters, RSF replicon, *lacI*, His_6-tag, S-tag	Merck Millipore (Billerica, MA, USA)
pUC18	Amp^R; pMB1 replicon, P*lac*, *lacI*, the 5'-terminal part of the *lacZ* gene encoding the N-terminal fragment of beta-galactosidase	Merck Millipore (Billerica, MA, USA)
pE-*Pcpal1*	pETDuet1 derivative containing *pal-1* from *P. crispum*	This study

pE-*Atpal2*	pETDuet1 derivative containing *pal-2* from *A. thaliana*	This study
pC-*Pdsts3-Sc4cl*	pCDFDuet1 derivative containing *sts-3* from *P. densiflora* and *4cl A294G* from *S. coelicolor*	This study
pC-*Pdsts3-At4cl2*	pCDFDuet1 derivative containing *sts-3* from *P. densiflora* and *4cl-2 N256A/M293P/K320L* from *A. thaliana*	This study
pC-*Pstrsts2-Sc4cl*	pCDFDuet1 derivative containing *sts-2* from *P. strobus* and *4cl A294G* from *S. coelicolor*	This study
pC-*Pstrsts2-At4cl2*	pCDFDuet1 derivative containing *sts-2* from *P. strobus* and *4cl-2 N256A/M293P/K320L* from *A. thaliana*	This study
pR-*Pstrsts2*	pRSFDuet1 derivative containing *sts-2* from *P. strobus*	This study
pR-*HisPdsts3*	pRSFDuet1 derivative containing *sts-3* from *P. densiflora* with a N-terminal His_6-tag	This study
pR-*HisPstrsts2*	pRSFDuet1 derivative containing *sts-2* from *P. strobus* with a N-terminal His_6-tag	This study
pE-*Pcpal1-Sc4cl*	pETDuet1 derivative containing *pal-1* from *P. crispum* and *4cl A294G* from *S. coelicolor*	This study
pE-*Pcpal1-At4cl2*	pETDuet1 derivative containing *pal-1* from *P. crispum* and *4cl-2 N256A/M293P/K320L* from *A. thaliana*	This study
pE-*Atpal2-Sc4cl*	pETDuet1 derivative containing *pal-2* from *A. thaliana* and *4cl A294G* from *S. coelicolor*	This study
pE-*Atpal2-At4cl2*	pETDuet1 derivative containing *pal-2* and *4cl-2 N256A/M293P/K320L* from *A. thaliana*	This study
pUC18-P*gap*-*HisPstrsts2-Sc4cl-Pcpal1*	pUC18 derivative containing *sts-2* from *P. strobus* with a N-terminal His_6-tag, *4cl A294G* from *S. coelicolor* and *pal-1* from *P. crispum* organized as operon under control of the promoter P*gap*	This study
pR-*HisPstrsts2-Sc4cl-Pcpal1*	pRSFDuet1 derivative containing *sts-2* from *P. strobus* with N-terminal His_6-tag, *4cl A294G* from *S. coelicolor* and *pal-1* from *P. crispum* organized as operon	This study
pR-T7-*accBC*-T7-*dtsR1*	pRSFDuet1 derivative containing *accBC* and *dtsR1,* encoding the biotin containing α-subunit	This study

	and core catalytic β-subunit, respectively, from the acetyl-CoA carboxylase (ACC) of *C. glutamicum*	
pE-*fabF*	pETDuet1 derivative containing β-ketoacyl-acyl carrier protein synthase II or FabF from *E. coli* BL21(DE3)	This study
pR-*HisPstrsts2T248A-Sc4cl-Pcpal1*	pRSFDuet1 derivative containing *sts-2 T248A* from *P. strobus* with N-terminal His_6-tag, *4cl A294G* from *S. coelicolor* and *pal-1* from *P. crispum* organized as operon	This study
pR-*HisPstrsts2Q361R-Sc4cl-Pcpal1*	pRSFDuet1 derivative containing *sts-2 Q361R* from *P. strobus* with N-terminal His_6-tag, *4cl A294G* from *S. coelicolor* and *pal-1* from *P. crispum* organized as operon	This study
pR-*HisPstrsts2T248A/Q361R -Sc4cl-Pcpal1*	pRSFDuet1 derivative containing *sts-2 T248A/Q361R* from *P. strobus* with N-terminal His_6-tag, *4cl A294G* from *S. coelicolor* and *pal-1* from *P. crispum* organized as operon	This study

engineering strategy for balancing the pathway with the *matB/matC* system from *Rhizobium trifolii*, which synthesizes the malonyl-CoA precursor from supplemented malonate. In comparison, reported product titers for pinosylvin are modest with 0.6 mg/L achieved with *Streptomyces venezuelae* (when supplementing 1.2 mM *trans*-cinnamic acid) (16) and 155 mg/L achieved during biotransformation with *E. coli* (when providing 1 mM *trans*-cinnamic acid and augmenting intracellular malonyl-CoA levels) (17).

In this study, we examined multiple genes and pathway configurations for the microbial pinosylvin production with *E. coli* from glucose. For this purpose, we optimized the heterologous gene expression, examined the precursor availability and performed laboratory protein evolution to enhance the activity of a plant-derived STS in the microbial host.

Materials and Methods

Bacterial strains, plasmids and growth conditions. Bacterial strains and plasmids used or constructed in the course of this work are listed in Table 1. *E. coli* DH5α, solely used for cloning purposes, was routinely cultivated on a rotary shaker (170 rpm) at 37°C in LB medium or on LB plates (LB medium with 1.5% [wt/vol] agar) (18). If appropriate, carbenicillin (50 µg/mL), kanamycin (30 µg/mL) or spectinomycin (50 µg/mL) was

TABLE 2 Oligonucleotides used in this study.

Oligonucleotide	Sequence (5' → 3') and properties[a]
Cloning of His_6-tagged Stilbene synthases	
F_BamHI_PdSTS3	GGATCCAATGGGTGGTGTTGATTTTGAAGGTTTTCG (*Bam*HI)
R_NotI_PdSTS3	GCGGCCGCTTATTACGGATGACAGGTAC (*Not*I)
F_BamHI_PstrSTS2	GGATCCAATGGGAAGCGTTGGTATGGGTG (*Bam*HI)
R_NotI_PstrSTS2	GCGGCCGCTTATTACGGAAACGGAATGC (*Not*I)
Construction of the pinosylvin operon under control of the T7 promoter	
F_Sc4CL_NotI	CGCAGTGCGGCCGCGAAAGGAGGTCTATATGTTTCGTAGCGAATATGC (*Not*I)
R_Sc4CL	CGCAGTGGTACCCGCATAGTACTTTATTAACGCGGTTCACGCAG (*Kpn*I/*Sca*I)
F_PcPAL1	CGCAGTGGTACCGAAAGGAGGTCTATATGGAAAATGGTAATGGTGCAAC (*Kpn*I)
R_PcPAL1_XhoI	CGCAGTCTCGAGTTATTAACAAATCGGCAGCGGTGCACCATTCCAG (*Xho*I)
Construction of the pinosylvin operon under control of the P_{gapdh} promoter	
F_P_{gapdh}_NdeI	CGCCATATGGATCAAACAGTGATATACGCCGTCAC (*Nde*I)
R_P_{gapdh}_HindIII	CGCAAGCTTATATTCCACCAGCTATTTGTTAGTGAATAAAAGG (*Hind*III)
F_HisPstrSTS2_SalI	CGCAGTGTCGACAAGGAGATATACCATGGGCAGCAGCCATCACCATC (*Sal*I)
R_HisPstrSTS2_NheI	CGCATGCTAGCTTATTACGGAAACGGAATGCTTTTCAGCACAAC (*Nhe*I)
F_Sc4CL	CGCAGTGGATCCGAAAGGAGGTCTATATGTTTCGTAGCGAATATGCAG (*Bam*HI)
R_Sc4CL	CGCAGTGGTACCCGCATAGTACTTTATTAACGCGGTTCACGCAG (*Kpn*I/*Sca*I)
F_PcPAL1	CGCAGTGGTACCGAAAGGAGGTCTATATGGAAAATGGTAATGGTGCAAC (*Kpn*I)
R_PcPAL1	CGCAGTGAATTCTTATTAACAAATCGGCAGCGGTG (*Eco*RI)
Construction of pR-T7-*accBC*-T7-*dtsR1*	

F_accBC_BamHI	CGCAGTGGATCCTGTCAGTCGAGACTAGGAAGATCACCAAG (*Bam*HI)
R_accBC_HindIII	CGCAGTAAGCTTTTATTACTTGATCTCGAGGAGAACAACGC (*Hind*III)
F_dtsR1_NdeI	CGCAGTCATATGATGACCATTTCCTCACCTTTGATTGACGTCG CCAACC (*Nde*I)
R_dtsR1_XhoI	CGCAGTCTCGAGTTATTACAGTGGCATGTTGCCGTGCTTG (*Xho*I)
Construction of pE-*fabF*	
F_fabF_NcoI	CGCAGTCCATGGGTGTCTAAGCGTCGTGTAGTTGTG (*Nco*I)
R_fabF_EcoRI	CGCAGTGAATTCGCAGCATGTTCACTACGGAACAAGTC (*Eco*RI)
PstrSTS3 error prone PCR	
F_PstrSTS2_Ep	CGCAGTGGATCCAATGGGAAGCGTTGGTATGGGTG (*Bam*HI)
R_PstrSTS2_Ep	CGCAGTGCGGCCGCTTATTACGGAAACGGAATGC (*Not*I)
Site directed mutagenesis at Q361 of PstrSTS3	
F_Q361R	CGTAAAGCAAGCCGTCGGAATGGTTGTAGCACC
R_Q361R	GGTGCTACAACCATTCCGACGGCTTGCTTTACG

[a] Recognition sites for the indicated restriction enzymes are underlined.

added to the medium. Growth was determined by measuring the optical density at 600 nm (OD_{600}).

Recombinant DNA techniques. The enzymes for recombinant DNA work were obtained from Thermo Scientific (Waltham, MA, USA) and Merck Millipore (Billerica, MA, USA). Routine methods like PCR, restriction, or ligation were carried out according to standard protocols (18). Oligonucleotides used for cloning, error prone PCR (epPCR) and site directed mutagenesis (19) were obtained from Eurofins Genomics (Ebersberg, Germany) and are listed in Table 2. *E. coli* was transformed by the RbCl method (20). The PAL, 4CL and STS genes were chemically synthesized by GeneArt Gene Synthesis services from Life Technologies, a Thermo Fisher Scientific company (Waltham, MA, USA). Genes were codon optimized for expression in *E. coli* employing the proprietary GeneOptimizer software. For subcloning purposes, all PAL and STS genes were synthesized with a 5'-*Nco*I and a 3'-*Bam*HI restriction site, whereas the synthetic 4CL genes incorporate a 5'-*Nde*I site and a 3'-*Xho*I site. The 5'-*Nde*I

and 5'-*Nco*I restriction sites already include the START-codon ATG as part of their recognition site. However, in case of the *sts-2* gene from *P. strobus*, subcloning using the *Nco*I site required the addition of the glycine codon GGA after the START-codon since the *Nco*I-recognition site (5'-CCATGG-3') comprises a G following the START-ATG codon.

Heterologous gene expression. Recombinant *E. coli* BL21(DE3) strains expressing genes for the pinosylvin synthetic pathway were cultivated in 25 mL LB medium in 100 mL baffled shake flasks at 37°C and 120 rpm until an OD_{600} of 0.6 was reached. After induction with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) the cultivation continued at 30°C for 24 hours. Cells were then harvested by centrifugation (3,900 g; 15 min; 4°C) and washed with 5 mL 0.9 wt% NaCl solution. After centrifugation (3,900 g; 15 min; 4°C) the cells were resuspended in 1 mL solution. After centrifugation (6,200 g; 10 min; 4°C), the cells were resuspended in 1 mL NNI10 lysis buffer (50 mM NaH_2PO_4; 300 mM NaCl; 10 mM imidazole; pH 8) and subsequently disrupted by sonication (intensity 2; duty cycle 20%; 4 min). The resulting CFE was clarified by centrifugation (16,000 g; 30 min; 4°C). A column volume of 1 mL nickel-nitrilotriacetic acid resin (Qiagen, Hilden, Germany) was used for the protein purification. After equilibration of the column by applying five column volumes (CV) of NNI10 buffer, the CFE was loaded phosphate-buffered saline (1.9 mM NaH_2PO_4; 8.1 mM Na_2HPO_4; 154 mM NaCl; pH 7.4) and disrupted by sonication using a Branson Sonifier 250 (intensity 2; duty cycle 20%; 4 min; Branson, Danbury, CT, USA). After removal of the cellular debris by centrifugation (16,000 g; 30 min; 4°C), the cell free extract (CFE) was subjected to SDS-PAGE analysis using NuPAGE Bis-Tris gels from Life Technologies (Waltham, MA, USA).

For the purification of the His_6-tagged STS proteins, *E. coli* BL21(DE3) cultures were cultivated in 100 mL LB medium in 500 mL baffled shake flasks on a rotary shaker (120 rpm) at 37°C until reaching an OD_{600} of 0.2. The cultures were shifted to 30°C and gene expression was induced with 0.1 mM IPTG as soon as an OD_{600} of 0.6 was reached. After four hours the cells were harvested by centrifugation (6,200 g; 10 min; 4°C) and washed with an ice cold 0.9 wt% NaCl on the column. Subsequently, the column was washed three times with 10 CV of NNI70 wash buffer (50 mM NaH_2PO_4; 300 mM NaCl; 70 mM imidazole; pH 8). The elution of the STS proteins was achieved by applying six times one CV of NNI250 elution buffer (50 mM NaH_2PO_4; 300 mM NaCl; 250 mM imidazole; pH 8). The eluted fractions were judged by SDS-PAGE analysis.

Furthermore, the successful expression of all PAL, 4CL and HisSTS proteins was verified by MALDI-ToF-MS as previously described (21).

Pinosylvin production with *E. coli*. Recombinant *E. coli* BL21(DE3) strains were either cultivated in 50 mL LB broth or in 50 mL Yeast Nitrogen Base (YNB) defined medium in 500 mL baffled shake flasks on a rotary shaker (120 rpm) at 26°C. For the preparation of 1 L YNB medium 100 mL of 10 times concentrated YNB was added to 900 mL base medium (6 g K_2HPO_4; 3 g KH_2PO_4; 10 g 3-(N-Morpholino)propanesulfonic acid (MOPS); pH 7) and 5 g/L glucose was added as carbon source. Precultures were cultivated overnight at 37°C (170 rpm) in LB or YNB medium, then diluted to an OD_{600} of 0.1 in 50 mL fresh LB or YNB medium, respectively. When the OD_{600} of these main cultures reached 0.6, gene expression was induced with 1 mM IPTG (when necessary) and the cultivation was continued for 36 hours at 26°C. Samples were taken at regular time intervals for LC-MS analysis to follow the formation of pinosylvin.

A two-stage cultivation procedure was also employed for pinosylvin production. In the first stage, cultivation and gene expression were performed in LB as described above. After 5 hours of heterologous gene expression, cells were collected by centrifugation (4,300 g; 15 min; 4°C) and subsequently resuspended in 10 mL YNB medium 5 g/L glucose and supplemented with 1 mM IPTG. The cultivation of this second phase was continued for 36 hours at 26°C. Again, samples were taken at regular time intervals for LC-MS analysis.

Metabolite extraction. Metabolite extracts from supernatant and cells were prepared for LC-MS analysis by mixing 1 mL culture with 1 mL ethyl acetate and vigorous shaking (1,400 rpm; 10 min; 20°C) in an Eppendorf thermomixer (Hamburg, Germany). 800 µl of this suspension were centrifuged for five minutes at 16,000 g and the ethyl acetate layer was transferred to an organic solvent resistant deep-well plate (Eppendorf, Hamburg, Germany). After evaporation of the ethyl acetate, 100 µl acetonitrile, concentrating the extract eight-fold, were added prior to LC-MS analysis.

Determination of cytoplasmic metabolite concentrations. *E. coli* cells were separated from the medium and inactivated by silicone oil centrifugation for the determination of cytoplasmic concentrations of *trans*-cinnamic acid, *trans*-cinnamoyl CoA, malonyl-CoA and pinosylvin as described previously (22). The resulting upper-phase was used for further analysis by LC-MS/MS.

Chemical synthesis of *trans*-cinnamoyl-CoA. *trans*-cinnamic acid was esterified in the presence of CoA and dicyclohexyl carbodiimide as previously described to prepare a standard for the LC-MS/MS analysis of *trans*-cinnamoyl-CoA (23). All required chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA).

Determination intermediates and pinosylvin by LC-MS and LC-MS/MS. *trans*-cinnamic acid and pinosylvin in cell-free culture medium as well as in the ethyl acetate extracted samples were quantified by LC-MS using an ultra-high-performance LC (uHPLC) 1290 Infinity system coupled to a 6130 Quadrupol LC-MS system (Agilent, Santa Clara, CA, USA). LC separation was carried out with a kinetex 1.7u C18 100A column (50 mm x 2.1 mm internal diameter; Phenomenex, Torrance, CA, USA) at 50°C. For elution, 0.1% acetic acid (solvent A) and acetonitrile supplemented with 0.1% acetic acid (solvent B) were applied as the mobile phases at a flow rate of 0.5 mL/min. A gradient was used, where solvent B was stepwise increased (minute 0 – 1: 15% to 22%; minute 1 – 2: 22% to 40%; minute 2 – 2.5: 40% to 50%; minute 2.5 – 3: 50% to 100%). The mass spectrometer was operated in the negative electrospray ionization (ESI) mode and data acquisition was performed in selected-ion-monitoring mode. LC-MS/MS was used for the quantification of *trans*-cinnamic acid, *trans*-cinnamoyl-CoA, malonyl-CoA and pinosylvin in the cytoplasmic extracts following a previously described method (24). *Trans*-cinnamic acid, malonyl-CoA and pinosylvin standards were purchased from Sigma-Aldrich (St. Louis, MO, USA).

Construction of *Pstrsts-2* epPCR libraries and microtiter plate screening for increased pinosylvin production. All epPCRs of the *sts-2* gene from *P. strobus*, with mutation rates varying from 2 - 4.6 mutations/kb, were performed using the Diversify PCR Random Mutagenesis Kit from Clontech (Mountain View, CA, USA). The mutated *Pstrsts2* genes were sub-cloned into the pRSFduet1 backbone carrying the genes for the PcPAL1 and the Sc4CL, (resulting in the vector pR-*HisPstrsts2*-Sc4cl-Pcpal1*) and transformed to *E. coli* BL21(DE3). At least 90 clones for each mutation rate, always accompanied by a negative control (*E. coli* BL21(DE3) pRSFDuet1) and a strain expressing the wild type *sts-2* gene (*E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1*) were screened for improved pinosylvin titers. For this purpose cultivation and screening were performed in the 96-well microtiter plate format. The libraries were cultivated in 600 µl YNB medium supplemented with 5 g/L glucose, 3 mM L-phenylalanine, 0.5 mM IPTG and 30 µg/mL kanamycin in deep-well plates from Eppendorf (Hamburg, Germany) in a microtron shaker from Infors (Bottmingen, Switzerland) at 30°C, 900 rpm and 75% humidity for 24 hours. Subsequently, 100 µl culture of each well was transferred to a UV transparent 96 well flat bottom plate from Corning (Corning, NY, USA). An Infinite M200 PRO microplate reader (Tecan, Männedorf, Switzerland) was used to relatively compare pinosylvin titers based on the fluorescent properties of pinosylvin (λ_{ex}: 316 nm; λ_{em}: 388 nm).

Results

Design of a synthetic pathway for pinosylvin. For the construction of a synthetic pinosylvin pathway we first consulted recent literature and the *BRaunschweig ENzyme DAtabase* (BRENDA) to identify suitable phenylalanine ammonia lyases (PALs), 4-coumarate-CoA ligases (4CLs) and stilbene synthases (STSs) originating either from plants or microorganisms (25). Enzymes were considered as suitable if they had been already functionally expressed in *E. coli* and if the determined basic catalytic parameters (e.g. k_{cat}, K_m) appeared to be promising for the construction of an efficient biosynthetic route towards pinosylvin. In the course of that survey, we selected the PAL isozyme 1 from *Petroselinum crispum* (PcPAL1) and the PAL isozyme 2 from *Arabidopsis thaliana* (AtPAL2) since both enzymes are characterized by a high activity when expressed in *E. coli*. Furthermore, we decided to express a variant of a 4CL-like enzyme (Sc4CL) from the gram-positive bacterium *Streptomyces coelicolor*, which favors *trans*-cinnamic acid over *p*-coumaric acid as substrate and a variant of the 4CL isozyme 2 of *Arabidopsis thaliana* (At4CL2). As candidates for stilbene synthases we selected the STS isozyme 3 of *Pinus densiflora* (PdSTS3) and STS isozyme 2 of *Pinus strobus* (PstrSTS2) as these isozymes were reported to have a high specificity and activity towards cinnamoyl-CoA.

Optimization of the heterologous gene expression. For initial gene expression experiments, all PAL genes were cloned into the pETDuet1-vector, whereas the STS- and 4CL-genes were cloned into pCDFduet1-vector under the individual control of a T7-promoter. The successful expression of soluble protein could be verified for all PALs and 4CLs by SDS-PAGE analysis. However, no visible bands indicating successful expression of PdSTS3 or PstrSTS2, could be observed in the soluble or the insoluble protein fractions (data not shown). Neither incremental reduction of the expression temperature from 30°C to 15°C, nor variation of the IPTG concentration (0.4 mM to 1 mM) improved STS expression. Furthermore, a first analysis of the supernatant of the expression cultures showed that up to 1.7 mM *trans*-cinnamic acid, but no pinosylvin was produced by the recombinant *E. coli* strains. In order to improve the expression of PdSTS3 and PstrSTS2, both genes were individually cloned into the pRSFDuet1 vector, another member of the pDuet vector family, with the RSF origin of replication allowing for higher copy numbers of the plasmids per cell (>100 copies per cell). In addition, both genes were cloned with a N-terminal His_6-tag, since such a tag improved the soluble expression of PdSTS3 without affecting the activity of the STS in *in vitro* enzyme assays (26). Despite these alterations, no soluble STS expression was observable in CFEs, but both His_6-

tagged enzymes (HisPstrSTS2 and HisPdSTS3) could be identified in the whole cell fraction. Purification of the enzymes from the CFE by affinity chromatography finally revealed low expression of soluble HisPstrSTS2 protein but not HisPdSTS3 in the CFE (Figure S1, supplemental material). Possibly, native *E. coli* proteins concealed the band corresponding to HisPstrSTS2 on the SDS-PAGE-analysis of the CFE. Finally, the identity of all PALs, 4CLs and HisSTS proteins was confirmed by MALDI-ToF-MS (data not shown). Accompanying chemical analysis of the cultures during the optimization of heterologous gene expression revealed that the modification of STS-expression already resulted in the accumulation of low concentrations of pinosylvin (0.068 ± 0.004 mg/L), but only in extracts of strains expressing the HisSTS of *P. strobus*.

Identification of the most suitable pathway configuration. After the identification of His_6-tagged PstrSTS2 as the only functional STS in *E. coli*, four different synthetic pathways with all combinations of the available PALs and 4CLs were constructed. Subsequently, pinosylvin production of all four strains was compared using the two-phase cultivation conditions as described in materials and methods. As it turned out, combination of the PAL from *P. crispum*, the 4CL from *S. coelicolor* and the STS from *P. strobus* proved to be the most productive pathway variant, but only a low product concentration of up to 0.64 mg/L pinosylvin could be detected (Table 3). Interestingly, the combination of PcPAL1 and At4CL2 did not lead to any detectable product formation although both enzymes functioned in other pathway configurations. The positive effect of the His_6-tag on PstrSTS2 expression was also reflected by increased pinosylvin titers as the same strain expressing PstrSTS2 without the His_6-tag, accumulated only up to 0.22 mg/L pinosylvin (Table 3).

At this point, all genes of the pinosylvin pathway were expressed under control of their own T7 promoter to rule out any polar expression effects. With the aim to evaluate the expression of the pathway from a single three-gene operon, two additional plasmids were constructed, in which the operon was either expressed under the control of the IPTG inducible T7 promoter (using the pRSFDuet1 vector backbone), or the constitutive *gap* promoter (P_{gap}) of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from *E. coli* (using the pUC18 vector backbone) (27). The *gap* promoter was selected, as it induces gene expression during cultivation on glucose as carbon source, thus allowing the continuous transcription of the three heterologous genes. Surprisingly, no pinosylvin formation could be detected when the synthetic operon was constitutively expressed during single phase cultivation in YNB supplemented with 5 g/L glucose. Only the accumulation of up to 10 mg/L *trans*-cinnamic acid

during the cultivation could be measured. In contrast, IPTG-induced expression from the T7 promoter of the tricistronic pRSFDuet1-construct led to a six fold increased pinosylvin titer with *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1* (up to 3.37 mg/L), when compared to the monocistronic organization of the same genes in *E. coli* BL21(DE3) pE-*Pcpal1-Sc4cl* pR-*HisPstrsts2* (Table 3).

Availability of precursors and intermediates during pinosylvin synthesis. At this stage, insufficient levels of the precursors L-phenylalanine and malonyl-CoA could be limiting for the overall product titers. In order to find out whether the availability of L-phenylalanine or malonyl-CoA were limiting, cultivations were performed in which 3 mM L-phenylalanine and/or cerulenin (up to 200 µM) were supplemented during the production phase (Fig. 2). Cerulenin is an antifungal antibiotic produced by *Cephalosporium caerulens*, which blocks fatty acid biosynthesis by inhibiting the β-ketoacyl-acyl carrier protein synthases FabB and FabF, thereby preventing channeling of malonyl-CoA into the pathway for fatty acid synthesis (28). The exclusive supplementation of L-phenylalanine had no positive effect on the pinosylvin production (3.29 ± 0.11 mg/L without supplementation vs. 3.49 ± 0.42 mg/L with supplementation), but resulted in the accumulation of the intermediate *trans*-cinnamic acid in the supernatant (data not shown). In contrast, addition of cerulenin drastically increased product titers up to 18-fold, reaching 59 mg/L pinosylvin at a concentration of 200 µM cerulenin (and no addition of L-phenylalanine).

Subsequently, cytoplasmic levels of malonyl-CoA in *E. coli* BL21(DE3) were determined to verify that the addition of cerulenin had an effect on the intracellular malonyl-CoA pool. These experiments revealed, that the intracellular malonyl-CoA level increased from 2 pmol/mg CDW (no cerulenin) to 105 pmol/mg CDW 1.5 hours after the addition of 200 µM cerulenin. In contrast, *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1* accumulates malonyl-CoA only up to 59 pmol/mg CDW in the presence of 200 µM cerulenin and

TABLE 3 Pinosylvin production with *E. coli* strains harboring different variants of the pinosylvin pathway in mono- or tricistronic transcriptional units.

Strain	Pinosylvin titer (mg/L) experiment 1	Pinosylvin titer (mg/L) experiment 2
E. coli BL21(DE3) pE-*Pcpal1-Sc4cl* pR-*HisPstrsts2*	0.49	0.64
E. coli BL21(DE3) pE-*Pcpal1-At4cl2* pR-*HisPstrsts2*	—[a]	—[a]
E. coli BL21(DE3) pE-*Pcpal1-Sc4cl* pR-*Pstrsts2*	0.21	0.22
E. coli BL21(DE3) pE-*Atpal2-Sc4cl* pR-*HisPstrsts2*	0.27	0.36
E. coli BL21(DE3) pE-*Atpal2-At4cl2* pR-*HisPstrsts2*	0.24	0.31
E. coli BL21(DE3) pUC18-P*gap-HisPstrsts2-Sc4cl-Pcpal1*	—[a]	—[a]
E. coli BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1*	3.21	3.37

[a] Below limit of detection

during pinosylvin synthesis (Figure S2, supplemental material), indicating malonyl-CoA consumption during pinosylvin formation.

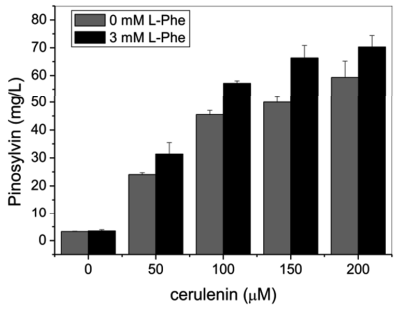


FIG 2 Pinosylvin production of *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1* with and without supplementation of L-phenylalanine and/or addition of cerulenin. The experiments were performed in triplicate.

L-phenylalanine supplementation had a positive impact on product formation during cerulenin-mediated inhibition of the fatty acid synthesis, indicating that the availability of this precursor becomes limiting when more malonyl-CoA is

available (Fig. 2). At this stage, supplementation of 3 mM L-phenylalanine and addition of 200 µM of cerulenin led to the highest pinosylvin titers of 70 mg/L after 36 hours of cultivation.

Two additional strategies were pursued for improving the intracellular malonyl-CoA availability in *E. coli* to circumvent cerulenin addition: Overexpression of the intrinsic β-ketoacyl-acyl carrier protein (ACP) synthase II (FabF) for reducing the drain of malonyl-CoA into fatty acid synthesis, as well as heterologous expression of the acetyl-CoA carboxylase from *Corynebacterium glutamicum* (ACC) for increasing the intracellular generation of malonyl-CoA from acetyl-CoA. Unfortunately heterologous expression of either enzyme only led to reduced pinosylvin titers (data not shown).

The availability of the intermediate *trans*-cinnamoyl-CoA could also be limiting during pinosylvin synthesis due to low 4CL-activity. Therefore, the relative intracellular levels of *trans*-cinnamic acid, *trans*-cinnamoyl-CoA and pinosylvin were determined during cultivation with 3 mM L-phenylalanine and with and without addition of cerulenin in the production phase (Fig. 3). During these experiments, L-phenylalanine was always supplemented to ensure that *trans*-cinnamic acid as direct precursor of *trans*-cinnamoyl-CoA is not limiting in the presence of cerulenin. All measurements were conducted two hours after addition of cerulenin, but only relative levels of intermediates and product could be determined, since no *trans*-cinnamoyl-CoA standard of sufficient purity could be chemically synthesized. As a result, addition of 200 µM of cerulenin led to an 18-fold increase of pinosylvin, accompanied by a drop of the *trans*-cinnamoyl-CoA level of only 30%. This observation hints at sufficient levels of *trans*-cinnamoyl-CoA to achieve higher product titers as the pool of this intermediate is not depleted or even significantly reduced.

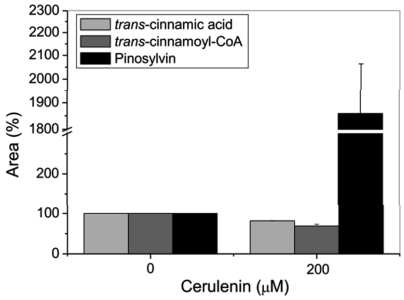


FIG 3 Relative intracellular levels of *trans*-cinnamic acid, *trans*-cinnamoyl-CoA and pinosylvin as determined by LC-MS/MS. The experiments with *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1* were performed in duplicate with supplementation of 3 mM L-phenylalanine and 0 µM or 200 µM cerulenin. The relative levels are given in "Area %" with the intermediates determined at 0 µM cerulenin set to 100%.

Directed protein evolution of HisPstrSTS2. The PAL- and 4CL-activity in *E. coli* appears to be sufficient to

support higher product titers as the intracellular pools of *trans*-cinnamic acid and *trans*-cinnamoyl-CoA are not depleted during pinosylvin production. In contrast, expression of the STS was still low although the gene sequences of all STS genes were optimized for an expression in *E. coli* prior to gene synthesis. Hence, the poor soluble expression of HisPstrSTS2 in the microbial host might be limiting the performance of the pinosylvin pathway. With the aim to adapt the STS2 of *P. strobus* for an optimal expression in *E. coli*, we performed directed evolution of the HisPstrSTS2 *in vivo*. For this purpose we randomly mutated the codon optimized open reading frame of *HisPstrsts2* by epPCR and subcloned the resulting STS-library into pR-*Sc4cl-Pcpal* to complete the pinosylvin pathway. Cultivation and screening of the library was performed in the MTP-format, taking advantage of the fluorescence of pinosylvin for identifying clones with increased product titers in comparison to the starting strain *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1*. Screening of 450 clones yielded two clones exhibiting a higher fluorescence under screening conditions. DNA sequencing revealed that each STS gene carries a single point mutation (A742G and A1082G), both leading to amino acid substitutions (T248A and Q361R, respectively). After recloning of the STS variants and retransformation to exclude any effect of the plasmid or strain background, the product titers of both variants were compared to the starting strain using the optimized cultivation conditions with and without the addition of cerulenin or the supplementation of L-phenylalanine.

Under all cultivation conditions tested, both amino substitutions in HisPstrSTS2 individually improved the overall product titers (Fig. 4). In a direct comparison, the positive effect of T248A for the pinosylvin formation was more pronounced as more pinosylvin could be detected in the extracts. In the presence of cerulenin, but without supplementation of L-phenylalanine, 70 mg/L pinosylvin accumulated in the *E. coli* BL21(DE3) pR-*HisPstrsts2T248A-Sc4cl-Pcpal1* cultures, whereas an additional supplementation of 3 mM L-phenylalanine increased the pinosylvin titer to 91 mg/L. However, the combination of both amino acid substitutions in an attempt to further boost product formation failed and most likely resulted in an inactive STS variant as no pinosylvin synthesis of *E. coli* BL21(DE3) pR-*HisPstrsts2T248A/Q361R-Sc4cl-Pcpal1* was detected.

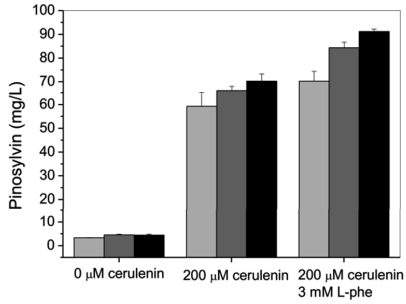


FIG 4 Comparison of the pinosylvin titers of *E. coli* BL21(DE3) pR-*HisPstrsts2-*

Sc4cl-Pcpal1 expressing either *HisPstrsts2*-wt (light grey), *HisPstrsts2*-Q361R (dark grey) or *HisPstrsts2*-T248A (black) without any supplementation, addition of 200 µM cerulenin or addition of 200 µM cerulenin and supplementation of 3 mM L-phenylalanine. The experiments were performed in duplicate.

Discussion

The development of efficient microbial platform organisms for the production of (plant) natural products requires the identification of suitable enzymes, the design of stable genetic constructs for gene expression and the optimal adaptation of the synthetic pathway to the host cell metabolism (or vice versa). In this work, we systematically explored several strategies to engineer *E. coli* towards the microbial production of the stilbene pinosylvin. The efforts included the optimization of gene expression conditions, the comparison of different expression constructs and protein engineering to adapt the expression of the plant-derived STS2 from *P. strobus* to the microbial host system.

Since construction of stilbene producing microorganisms has been mostly limited to strains requiring the supplementation of phenylpropanoid intermediates, we decided to design and implement the complete three-step pathway to pinosylvin starting from the amino acid L-phenylalanine. We set off with the identification of the most suitable enzymes from various microbial or plant sources in literature and databases, and selected two enzymes for each of the three required enzymatic steps based on their kinetic properties. PALs, catalyzing the first committed step towards phenylpropanoid synthesis, can be found in all higher plants, but are also present in yeast, fungi and some prokaryotic species (29). The PAL isozyme 1 from *Petroselinum crispum* (PcPAL1) and the PAL isozyme 2 from *Arabidopsis thaliana* (AtPAL2) were selected for the construction of the pinosylvin pathway since both enzymes are characterized by a high activity when expressed in *E. coli* (30, 31). In comparison to PALs, 4CLs are specific to the secondary metabolism of plants. Most 4CLs described, are substrate specific since all described natural substrates are characterized by a 4´-hydroxyl group on the phenyl ring (32). However, a synthetic pathway leading to pinosylvin requires a CoA-ligase that accepts *trans*-cinnamic acid lacking this *para*-hydroxy moiety. Interestingly, Kaneko and coworkers discovered a bacterial 4CL-like enzyme from the filamentous, soil-dwelling, gram-positive bacterium *Streptomyces coelicolor* (33). The enzyme showed a distinct 4CL activity, favoring *trans*-cinnamic acid over *p*-coumaric acid as substrate. Mutational analysis of the substrate binding pocket identified the amino acid substitution A294G conferring an even higher catalytic activity towards *trans*-cinnamic acid compared to the

native enzyme. This particular enzyme variant (Sc4CL) was selected for an evaluation in the synthetic pinosylvin pathway. The second 4CL candidate was a variant of the 4CL isozyme 2 of *Arabidopsis thaliana* (At4CL2) bearing three amino acid substitutions (N256A, M293P, K320L). In *in vitro* enzyme assays, this mutein demonstrated a 30-fold higher conversion of *trans*-cinnamic acid compared to the native enzyme (34). Pinosylvin-forming stilbene synthases (EC 2.3.1.146), are typical for gymnosperms and to date several such enzymes have been identified in three pine species (*Pinus sylvestris* (35), *Pinus densiflora* (26) and *Pinus strobus* (36)). STS isozyme 3 of *P. densiflora* (PdSTS3) and STS isozyme 2 of *P. strobus* (PstrSTS2) were selected as STS candidates, as these isozymes were reported to have higher specificity and activity towards cinnamoyl-CoA in comparison to the other isozymes in *in vitro* enzyme assays (26, 36). Interestingly, compared to other characterized STSs, PdSTS3 has a truncated C-terminus, which is believed to be responsible for the observed release from feedback inhibition by pinosylvin (26).

The combination of PcPAL1 and Sc4CL in the synthetic pinosylvin pathway turned out to be more beneficial for high product titers than the combination of AtPAL2 and Sc4CL. This is interesting because the At4CL2 performed better with the AtPAL2 than with the PAL from *P. crispum*. This could be simply due to the common origin of both enzymes from *A. thaliana* but also supports the notion that the performance of an entire synthetic pathway should be evaluated instead of just assembling a pathway from individually characterized “best parts”. Pinosylvin production could be improved by organizing the three genes of the final pinosylvin pathway configuration in an operon under the control of a single promoter. This same effect was also observed during the development of a microbial production strain for resveratrol, where cotranscription of the genes for 4CLs and STSs was believed to lead to a more balanced gene expression (11).

The malonyl-CoA availability turned out to be crucial for the overall product titers as the addition of the fatty acid production inhibitor cerulenin boosted the pinosylvin titers 18 to 20-fold, whereas the supplementation of L-phenylalanine alone had no effect. In order to avoid the addition of cerulenin, two strategies for increasing the intracellular malonyl-CoA availability were followed: overexpression of *E. coli*´s own FabF (37) as well as heterologous expression of the ACC from *C. glutamicum* as it was successfully demonstrated during microbial polyketide synthesis and flavanone synthesis, respectively (17, 38, 39). However, none of these approaches had the desired effect as the pinosylvin production significantly decreased or ceased entirely. No improvement of the microbial flavonoid production was also observed during

microbial flavonoid production with *E. coli* upon expression of the ACC from *C. glutamicum*, suggesting that this strategy is not generally applicable (40). A comparative analysis of intracellular intermediate availability and product concentration in the presence or absence of 200 µM cerulenin was conducted, but the obtained results did not reveal a limiting enzymatic step when more malonyl-CoA is available. Additional experiments will be necessary to shed more light on this subject.

Finally, we performed directed protein evolution to adapt the expression of the pine-derived HisPstrSTS2 to the microbial host system. This was directly done in the genetic context of the pinosylvin pathway. Only one round of diversity generation by epPCR and MTP-based screening of 450 clones for increased pinosylvin-related fluorescence identified two clones with improved product formation, each bearing a single amino acid substitution. The combination of both mutations had a detrimental effect on product formation, but the T248A substitution alone increased the final pinosylvin titer of the best strain *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1* from 59 mg/L to 70 mg/L in the presence of 200 µM cerulenin, and by an additional 30% to 91 mg/L when L-phenylalanine was also supplemented. With the aim to understand the structural consequences of both amino acid substitutions, we generated a structure model for a PstrSTS2-dimer using SWISS-MODEL (41). A structure model for the pinosylvin synthase of *P. sylvestris* (PDB code: 1U0U, (42)), which was not among the pinosylvin synthases tested by us, served as template for the calculated model since this enzyme shares 87% sequence identity to PstrSTS2 on the protein level. A structural alignment with structure models from the closely related chalcone synthase from *Medicago sativa* with coenzyme A bound (PDB code: 1BQ6 (43)) and the stilbene synthase from *Arachis hypogaea* with resveratrol bound (PDB code: 1Z1F (44)) helped us to pinpoint the active site in the calculated PstrSTS2 model. Both substitutions are not in close proximity to the assumed active sites of the PstrSTS2 dimer (Fig. 5A). T248 is located in the middle of a long β-strand and its side chain is facing the interior of the protein at the dimer interface. The substitution by an L-alanine at this position could simply improve the flexibility of the protein and might ultimately lead to the observed improvement of the STS activity in *E. coli.* In contrast, Q361 is positioned at the end of an α-helix on the protein surface and located next to the amino acid sequence N362-G363-C364. This pattern follows the typical *N*-glycosylation sequence N-X-S/T/C of L-asparagine (*N*)-linked protein glycosylation, where “X” can be any amino acid except L-proline (45, 46). Analysis of the PstrSTS2 amino acid sequence with the web server “*GlycoEP”* also identified N362 as potential site for *N*-glycosylation

in this particular STS (47). *E. coli* cannot perform L-asparagine (*N*)-linked protein glycosylation. Hence, substitution of L-glutamine with the more hydrophilic L-arginine at position 361 might simply lead to an improved stability or solubility of PstrSTS2 in the absence of any glycosylation in the heterologous host. When both amino acid substitutions were

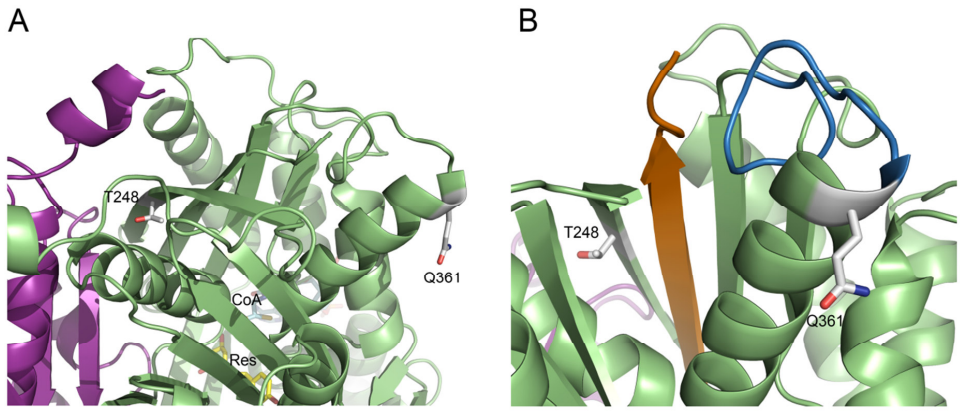


FIG 5 (A) Cartoon representation of a close-up of the pinosylvin synthase STS2 from *P. strobus*. The structure model was calculated using the model of the pinosylvin synthase of *P. sylvestris* as template. The monomers are shown in green and purple, resveratrol (Res) and coenzyme A (CoA) highlighted in one monomer are shown in the stick mode in yellow and cyan/orange, respectively. T248 is located at the dimer interface, whereas Q361 is located on the protein surface. Both amino acids are shown in the stick mode. (B) T248 and Q361 both interact with the C-terminal β-strand shown in orange. The β-strand, in which T248 is located, is adjacent to this C-terminal β-strand and the α-helix of Q361 interacts indirectly with the same C-terminal β-strand via a loop.

combined, no pinosylvin synthesis could be observed, suggesting that this enzyme variant is inactive. The β-strand, in which T248 is located, is adjacent to the C-terminal β-strand of PstrSTS2 and the α-helix of Q361 interacts indirectly with the same C-terminal β-strand via a loop (Fig. 5B). Possibly, the combination of both mutations simply destabilizes the architecture of PstrSTS2 rendering this enzyme variant inactive.

Key to success for the microbial production of pinosylvin from glucose without supplementation of any phenylpropanoid intermediate was the combination of various metabolic engineering approaches and protein engineering of a heterologously expressed key enzyme. However, at this point the addition of cerulenin is still required for achieving significant product titers. This could be circumvented by carrying out

multiple genetic alterations of the central carbon metabolism to redirect the carbon flux to malonyl-CoA as it has been done for the microbial synthesis of the flavanone naringenin with *E. coli* (48). Here, the successful modifications suggested by the Optforce model included the overexpression of the genes for the pyruvate dehydrogenase multi-enzyme complex, phosphoglycerate kinase, acetyl-CoA carboxylase and deletion of the genes encoding for the fumarase and succinyl-CoA synthetase. Following this strategy, 5.6 times higher naringenin titers could be attained.

Nonetheless, protein engineering by directed evolution in the genetic context of the synthetic pathway as presented in this manuscript could also be used to "harmonize" the entire pinosylvin pathway with *E. coli* for achieving higher product titers. This strategy might also be useful during the development of other microbial strains for the production of small metabolites of high value.

Acknowledgments

This work was supported by the BOOST-fund "PNP-EXPRESS" from the "NRW Strategieprojekt BioSC" and the ''Helmholtz Initiative Synthetic Biology'' from the Helmholtz Association.

We thank Petra Geilenkirchen (Forschungszentrum Jülich) for help with the LC-MS/MS analyses, Katharina Neufeld (University of Düsseldorf) for assisting in the chemical synthesis of *trans*-cinnamoyl-CoA and Marco Bocola (RWTH Aachen University) for useful advice with the structural interpretation of the obtained amino acid substitutions in the HisPstrSTS2 variants.

References

1. **Marienhagen J, Bott M.** 2013. Metabolic engineering of microorganisms for the synthesis of plant natural products. J. Biotechnol. **163:**166-178.
2. **Rivière C, Pawlus AD, Mérillon JM.** 2012. Natural stilbenoids: distribution in the plant kingdom and chemotaxonomic interest in Vitaceae. Nat. Prod. Rep. **29:**1317-1333.
3. **Chong JL, Poutaraud A, Hugueney P.** 2009. Metabolism and roles of stilbenes in plants. Plant Sci. **177:**143-155.
4. **Park EJ, Chung HJ, Park HJ, Kim GD, Ahn YH, Lee SK.** 2013. Suppression of Src/ERK and GSK-3/beta-catenin signaling by pinosylvin inhibits the growth of human colorectal cancer cells. Food Chem. Toxicol. **55:**424-433.
5. **Park EJ, Park HJ, Chung HJ, Shin Y, Min HY, Hong JY, Kang YJ, Ahn YH, Pyee JH, Lee SK.** 2012. Antimetastatic activity of pinosylvin, a natural stilbenoid, is associated with the suppression of matrix metalloproteinases. J. Nutr. Biochem. **23:**946-952.
6. **Jeong E, Lee HR, Pyee J, Park H.** 2013. Pinosylvin induces cell survival, migration and anti-adhesiveness of

endothelial cells via nitric oxide production. Phytother. Res. **27:**610-617.

7. **Jančinová V, Perečko T, Nosál' R, Harmatha J, Šmidrkal J, Drábiková K.** 2012. The natural stilbenoid pinosylvin and activated neutrophils: effects on oxidative burst, protein kinase C, apoptosis and efficiency in adjuvant arthritis. Acta Pharmacol. Sin. **33:**1285-1292.
8. **van Summeren-Wesenhagen PV, Marienhagen J.** 2013. Putting bugs to the blush: Metabolic engineering for phenylpropanoid-derived products in microorganisms. Bioengineered **4:**355-362.
9. **Conde E, Fang WW, Hemming J, Willför S, Domínguez H, Parajó JC.** 2014. Recovery of bioactive compounds from *Pinus pinaster* wood by consecutive extraction stages. Wood Sci. Technol. **48:**311-323.
10. **Ferré-Filmon K, Delaude L, Demonceau A, Noels AF.** 2004. Catalytic methods for the synthesis of stilbenes with an emphasis on their phytoalexins. Coord. Chem. Rev. **248:**2323-2336.
11. **Lim CG, Fowler ZL, Hueller T, Schaffer S, Koffas MAG.** 2011. High-yield resveratrol production in engineered *Escherichia coli*. Appl. Environ. Microbiol. **77:**3451-3460.
12. **Watts KT, Lee PC, Schmidt-Dannert C.** 2006. Biosynthesis of plant-specific stilbene polyketides in metabolically engineered *Escherichia coli*. Bmc Biotechnology **6**:22.
13. **Donnez D, Jeandet P, Clément C, Courot E.** 2009. Bioproduction of resveratrol and stilbene derivatives by plant cells and microorganisms. Trends Biotechnol. **27:**706-713.
14. **Bhan N, Xu P, Khalidi O, Koffas MAG.** 2013. Redirecting carbon flux into malonyl-CoA to improve resveratrol titers: Proof of concept for genetic interventions predicted by OptForce computational framework. Chem. Eng. Sci. **103:**109-114.
15. **Wu JJ, Liu PR, Fan YM, Bao H, Du GC, Zhou JW, Chen J.** 2013. Multivariate modular metabolic engineering of *Escherichia coli* to produce resveratrol from L-tyrosine. J. Biotechnol **167:**404-411.
16. **Park SR, Yoon JA, Paik JH, Park JW, Jung WS, Ban YH, Kim EJ, Yoo YJ, Han AR, Yoon YJ.** 2009. Engineering of plant-specific phenylpropanoids biosynthesis in *Streptomyces venezuelae*. J. Biotechnol **141:**181-188.
17. **Katsuyama Y, Funa N, Miyahisa I, Horinouchi S.** 2007. Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in *Escherichia coli*. Chem. Biol. **14:**613-621.
18. **Sambrook J, Russell D.** 2001. Molecular cloning. A laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. **Bauer JC, Wright DA, Braman JC, Geha RS.** 1998. Circular site-directed mutagenesis. US patent 5,789,166A.
20. **Hanahan D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:**557-580.

21. **Marienhagen J, Kennerknecht N, Sahm H, Eggeling L.** 2005. Functional analysis of all aminotransferase proteins inferred from the genome sequence of *Corynebacterium glutamicum*. J. Bacteriol. **187:**7639-7646.
22. **Marienhagen J, Eggeling L.** 2008. Metabolic function of *Corynebacterium glutamicum* aminotransferases AlaT and AvtA and impact on L-valine production. Appl. Environ. Microbiol. **74:**7457-7462.
23. **Gross GG, Zenk MH.** 1974. Isolation and properties of hydroxycinnamate:CoA ligase from lignifying tissue of *Forsythia*. Eur. J. Biochem. **42:**453-459.
24. **Paczia N, Nilgen A, Lehmann T, Gätgens J, Wiechert W, Noack S.** 2012. Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. Microb. Cell. Fact. **11:**122.
25. **Schomburg I, Chang A, Schomburg D.** 2014. Standardization in enzymology—Data integration in the world's enzyme information system BRENDA. Perspectives in Science **1:**15-23.
26. **Kodan A, Kuroda H, Sakai F.** 2002. A stilbene synthase from Japanese red pine (*Pinus densiflora*): Implications for phytoalexin accumulation and down-regulation of flavonoid biosynthesis. PNAS **99:**3335-3339.
27. **Charpentier B, Bardey V, Robas N, Branlant C.** 1998. The EII^{Glc} protein is involved in glucose-mediated activation of *Escherichia coli gapA* and *gapB-pgk* transcription. J. Bacteriol. **180:**6476-6483.
28. **Moche M, Schneider G, Edwards P, Dehesh K, Lindqvist Y.** 1999. Structure of the complex between the antibiotic cerulenin and its target, beta-ketoacyl-acyl carrier protein synthase. J. Biol. Chem. **274:**6031-6034.
29. **MacDonald MJ, D'Cunha GB.** 2007. A modern view of phenylalanine ammonia lyase. Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire **85:**273-282.
30. **Röther D, Poppe L, Morlock G, Viergutz S, Rétey J.** 2002. An active site homology model of phenylalanine ammonia-lyase from *Petroselinum crispum*. Eur. J. Biochem. **269:**3065-3075.
31. **McKenna R, Nielsen DR.** 2011. Styrene biosynthesis from glucose by engineered *E. coli*. Metab. Eng. **13:**544-554.
32. **Cukovic D, Ehlting J, VanZiffle JA, Douglas CJ.** 2001. Structure and evolution of 4-coumarate : coenzyme A ligase (4CL) gene families. Biol. Chem. **382:**645-654.
33. **Kaneko M, Ohnishi Y, Horinouchi S.** 2003. Cinnamate : coenzyme A ligase from the filamentous bacterium *Streptomyces coelicolor* A3(2). J. Bacteriol. **185:**20-27.
34. **Schneider K, Hövel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible HP.** 2003. The substrate specificity-determining amino acid code of 4-coumarate : CoA ligase. PNAS **100:**8601-8606.
35. **Schanz S, Schröder G, Schröder J.** 1992. Stilbene synthase from Scots

pine (*Pinus sylvestris*). FEBS Letters **313:**71-74.

36. **Raiber S, Schröder G, Schröder J.** 1995. Molecular and enzymatic characterization of 2 stilbene synthases from eastern white pine (*Pinus strobus*) - A single Arg/His difference determines the activity and the pH dependence of the enzymes. FEBS Letters **361:**299-302.
37. **Subrahmanyam S, Cronan JE.** 1998. Overproduction of a functional fatty acid biosynthetic enzyme blocks fatty acid synthesis in *Escherichia coli*. J. Bacteriol. **180:**4596-4602.
38. **Miyahisa I, Kaneko M, Funa N, Kawasaki H, Kojima H, Ohnishi Y, Horinouchi S.** 2005. Efficient production of (2S)-flavanones by *Escherichia coli* containing an artificial biosynthetic gene cluster. Appl. Microbiol. Biotechnol. **68:**498-504.
39. **Zha WJ, Rubin-Pitel SB, Shao ZY, Zhao HM.** 2009. Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. Metab. Eng. **11:**192-198.
40. **Leonard E, Lim KH, Saw PN, Koffas MAG.** 2007. Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. Appl. Environ. Microbiol. **73:**3877-3886.
41. **Arnold K, Bordoli L, Kopp J, Schwede T.** 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22:**195-201.
42. **Austin MB, Bowman ME, Ferrer JL, Schröder J, Noel JP.** 2004. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. Chem. Biol. **11:**1179-1194.
43. **Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP.** 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat. Struct. Biol. **6:**775-784.
44. **Shomura Y, Torayama I, Suh DY, Xiang T, Kita A, Sankawa U, Miki K.** 2005. Crystal structure of stilbene synthase from *Arachis hypogaea*. Proteins **60:**803-806.
45. **Matsui T, Takita E, Sato T, Kinjo S, Aizawa M, Sugiura Y, Hamabata T, Sawada K, Kato K.** 2011. N-glycosylation at noncanonical Asn-X-Cys sequences in plant cells. Glycobiology **21:**994-999.
46. **Strasser R.** 2014. Biological significance of complex N-glycans in plants and their impact on plant physiology. Front. Plant Sci. **5:**363.
47. **Chauhan JS, Rao A, Raghava GPS.** 2013. *In silico* Platform for Prediction of N-, O- and C-Glycosites in Eukaryotic Protein Sequences. PLoS One **8:** e67008.
48. **Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MAG.** 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab. Eng. **13:**578-587.

3.3

Combinatorial optimization of synthetic operons for the microbial production of *p*-coumaryl alcohol with *Escherichia coli*

Philana V. van Summeren-Wesenhagen[a], Raphael Voges[a], Alexander Dennig[b], Sascha Sokolowsky[a], Stephan Noack[a], Ulrich Schwaneberg[b] and Jan Marienhagen[a]*

[a] *Institut für Bio- und Geowissenschaften, IBG-1: Biotechnologie, Forschungszentrum Jülich, D-52425 Jülich, Germany*

[b] *Lehrstuhl für Biotechnologie, RWTH Aachen University, Worringer Weg 1, D-52056 Aachen, Germany*

Microbes are extensively engineered to produce compounds of biotechnological or pharmaceutical interest. However, functional integration of synthetic pathways into the respective host cell metabolism and optimization of heterologous gene expression for achieving high product titers is still a challenging task. In this manuscript we describe the optimization of a tetracistronic operon for the microbial production of the plant-derived phenylpropanoid *p*-coumaryl alcohol in *E. coli*. This was achieved by advancing the phosphorothioate-based ligase-independent gene cloning method for the rapid combinatorial assembly of synthetic operons. Operon libraries constructed with this method offer the opportunity for simultaneously balancing of gene expression on the level of translation to maximize product titers. This method is sequence independent, enzyme free, allows for easy automation due to its simplicity and modularity.

Tools and methods of synthetic biology have become increasingly important to design, assemble and optimize biosynthetic pathways for the biotechnological production of fine chemicals (*1*), biofuels (*2*), pharmaceuticals (*3*) or natural products *(4)*.

Tuning of gene expression from such synthetic pathways can be achieved on the level of transcription or translation. Transcription in particular has been in the focus of metabolic engineering since many years, as a myriad of different (inducible) natural promoters are known and have been successfully used for the tuning of gene expression. Furthermore, engineering of the promoter architecture has achieved tunable gene expression from synthetic promoters in various

industrially relevant platform organisms such as *Escherichia coli* (*5*) or *Saccharomyces cerevisiae (6)*. Gene expression on the translational level can be controlled via adaptation of the codon usage (*7*) during the design of the synthetic genes, modulation of the mRNA-stability (*8*), application of metabolite-responsive riboswitches (*9*) or design of the ribosomal binding site (RBS) (*10*). However, considering the complexity of gene expression, a balanced expression of multigene pathways for achieving optimal catalytic efficiencies still represents a major challenge, especially if genes of the respective pathways originate from different organisms. Combinatorial assembly and subsequent evaluation of many different pathway variants can be a very time consuming and expensive task depending on the cloning methods employed. Many DNA assembly techniques have been developed over the past few years to complement the more traditional cloning, which is based on the use of enzymes for *in vitro* DNA restriction and ligation (*11*). Among the more prominent methods for the assembly of pathways are Polymerase cycling assembly (PCA) (*12*), Gibson assembly (*13*), the Sequence and ligation independent cloning (SLIC) method (*14*) or USER-fusion (*15*).

The Phosphorothioate-based ligase-independent gene cloning (PLICing) method has been developed as an enzyme-free and sequence-independent cloning method (*16*) of single genes. PLICing starts with amplification of the target gene and the vector by PCR using primers with complementary phosphorothioated nucleotides at the 5´-end. The PCR products are cleaved in an iodine/ethanol solution, yielding single-stranded overhangs. Subsequently, these ends hybridize at room temperature and the resulting DNA constructs can be directly transformed into competent host cells. PLICing is sequence independent, and has found numerous applications in the field of protein engineering for performing simultaneous site-saturation of five codons in single genes (*17, 18*) and for recombining secondary structure elements, motifs or domains of single proteins (*19*).

In this study, we advanced the PLICing method for the rapid assembly of synthetic pathways and included the possibility to simultaneously tune ("balance") the expression of all pathway genes on the level of translation. This balancing is required for maximizing product titers and for minimizing inclusion bodies formation and potentially cytotoxic effects of accumulating pathways intermediates.

Model pathway for the development of this method was a four-step pathway to convert the amino acid L-tyrosine to the monolignol *p*-coumaryl alcohol (Figure 1). This plant natural product is an important precursor of pharmaceutically interesting lignans and

key building block of the plant polymer lignin. *p*-coumaryl alcohol synthesis from L-tyrosine starts with a deamination step catalyzed by a tyrosine ammonia lyase (TAL, EC 4.3.1.23) to form *p*-coumaric acid. Subsequently, this acid is coenzyme A (CoA)-activated by a 4-coumarate-CoA-ligase (4CL, EC 6.2.1.12). The resulting *p*-coumaryl-CoA is reduced to *p*-coumaryl alcohol in two subsequent steps by a cinnamoyl-CoA-reductase (CCR, 1.2.1.44) and a cinnamyl alcohol dehydrogenase (CAD, EC 1.1.1.195). The heterologous production of *p*-coumaryl alcohol was recently established in *Escherichia coli*, but without precursor feeding a final product titer of only 22 mg/L *p*-coumaryl alcohol could be achieved (*20*). This was explained by the observed cytoplasmatic aggregation of insoluble heterologous proteins in inclusion bodies.

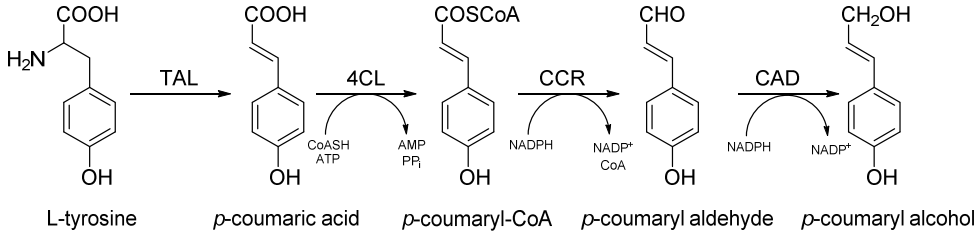


Figure 1. Biosynthetic pathway for the production of *p*-coumaryl alcohol from L-tyrosine.

At the beginning of our experiments, four genes encoding the tyrosine ammonia lyase from *Rhodobacter sphaeroides* (RsTAL), the 4-coumarate-CoA-ligase from *Petroselinum crispum* (Pc4CL), the cinnamoyl-CoA-reductase from *Zea mays* (ZmCCR) and the cinnamyl alcohol dehydrogenase from *Zea mays* (ZmCAD) were codon optimized for the heterologous expression in *E. coli* and synthesized. Initially, three plasmid-based variants of a tetracistronic *p*-coumaryl alcohol pathway were constructed. In all cases the transcription was controlled by a single IPTG-inducible T7 promoter and the translation of each gene in the pathways was modulated by an individual, but always identical Shine-Dalgarno (SD) sequence (5´-TAAGGAGGT-3´) as prokaryotic ribosome binding site (pALXtreme-1a_T7$_{prom}$>*Rstal-Pc4cl-Zmccr-Zmcad*). The pathways differed only in the length (spacing) of the nucleotide sequence between SD-sequence and the START codon of each gene. The length of this short spacing is known to have major influence on the translation initiation efficiency (*21*). A spacing of 5 nucleotides (nt) between these two regulatory elements confers the optimal translational efficiency, whereas a deviation from this length (either a reduction or an elongation) results in stepwise decreased translation efficiency. In the first pathway variant (5-5-

5-5 variant) this spacing was set to 5 nt for each gene, in the second pathway (9-9-9-9 variant) to 9 nt and in the third pathway (13-13-13-13 variant) to 13 nt with the aim to generate three reference strains with a high, medium and low expression of the four synthetic *p*-coumaryl-alcohol pathway genes.

Subsequently, relative cytoplasmatic protein concentrations in cell free extract (CFE) were determined to verify that the variation of the spacing between the SD sequence and the START codon can indeed modulate the translation efficiency for each heterologously expressed protein. This was achieved by targeted proteomics applying isotope dilution mass spectrometry coupled to high performance liquid chromatography (IDMS-LC-MS/MS) (*22*). This method using ^{15}N-metabolic labeling of the target proteins as internal standards allows reliable peptide and protein quantification in CFEs as complex biological samples without the need for a complete separation of the respective analytes. For this purpose, four proteotypic peptides, generated by trypsin-digestion of CFEs, had to be identified by LC-MS/MS for each enzyme to serve as signature peptides during relative protein quantifications. The ratio of the areas for these peptides as determined by LC-MS/MS for each of the four enzymes in comparison to the same peptides in the ^{15}N-labeled form (added to the samples as internal standards), was used for relative protein quantification. The results obtained for the SD sequence – START-codon spacing of 5 nt was set to 1 for comparison of the obtained results (Figure 2). As expected, the highest cytoplasmatic protein concentrations were obtained with a spacing of 5 nt for all four heterologously expressed genes. With an increasing length of this spacing the relative cytoplasmic amount of all four proteins decreased stepwise. However, to a different degree for each protein. For instance, RsTAL expression was reduced by 75% when increasing the SD sequence-START codon spacing from 5 nt to 9 nt, but the amount of the Pc4CL was only reduced by 10%.

All three strains were subsequently characterized regarding their capability to produce *p*-coumaryl alcohol to find out if the observed decreased translation efficiency is also reflected by the product titers. The obtained results showed indeed, that the engineered variation of the SD sequence - START-codon spacing modulates overall product titers: The 5-5-5-5 variant accumulated up to 40 mg/L *p*-coumaryl alcohol, whereas a product concentration of 12 mg/l could be determined in the supernatant of the 9-9-9-9 variant. For the 13-13-13-13 variant the lowest *p*-coumaryl alcohol titer of only 5 mg/L could be measured.

Encouraged by these results we constructed the first combinatorial *p*-coumaryl-alcohol operon library with a

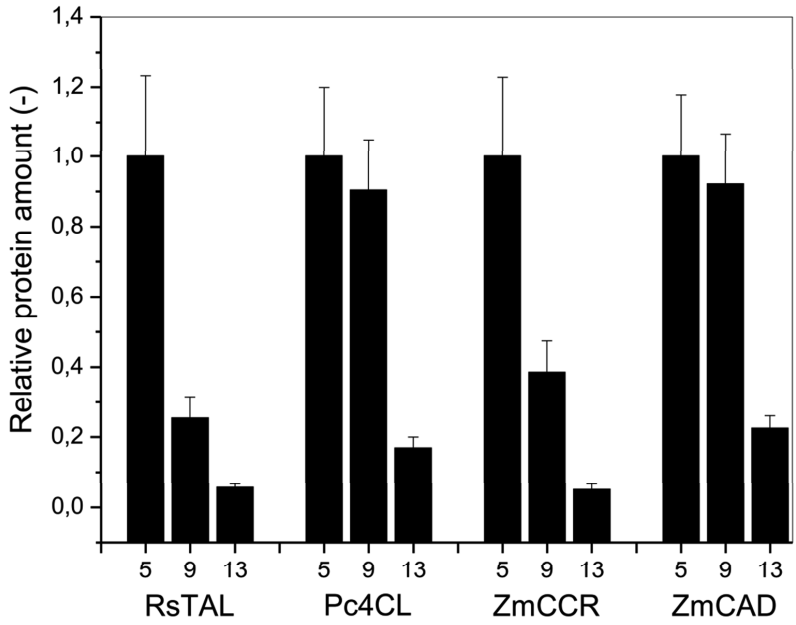


Figure 2. Relative cytoplasmatic protein concentrations of RsTAL, Pc4CL, ZmCCR and ZmCAD when heterologously expressed in *E. coli* BL21 (DE3) *lacI*^{Q1}. For each protein, three gene variants with different SD sequence - START codon spacings (5, 9, or 13 nt) were expressed. Relative protein concentrations were determined by IDMS-LC-MS/MS, and results for each protein, whose gene had a 5 nt SD sequence - START codon spacing upstream of the open reading frame was set to 1 for comparison.

variation of the translation efficiency for each gene in the four-step pathway following the PLICing principle (Figure 3). For this purpose, three versions of each gene, having either a 5, 9 or 13 nt SD sequence - START codon spacing were generated by PCR with phosphorothioate-oligonucleotides (Table S1). The twelve resulting PCR products and the pALXtreme backbone, which was also PCR amplified with phosphorothioate-oligonucleotides were individually subjected to chemical cleavage of the phosphorothioate bonds for generating complementary 3´-overhangs. Key to the combinatorial assembly of the pathway variants was the design of the complementary sequences required for hybridization of the five individual fragments, which make up a full operon. In this experiment, each of the required five sequences was unique to keep the gene order (*Rstal-Pc4cl-Zmccr-Zmcad*) of the synthetic *p*-coumaryl-alcohol operon during the process of random assembly constant. This limited the number of

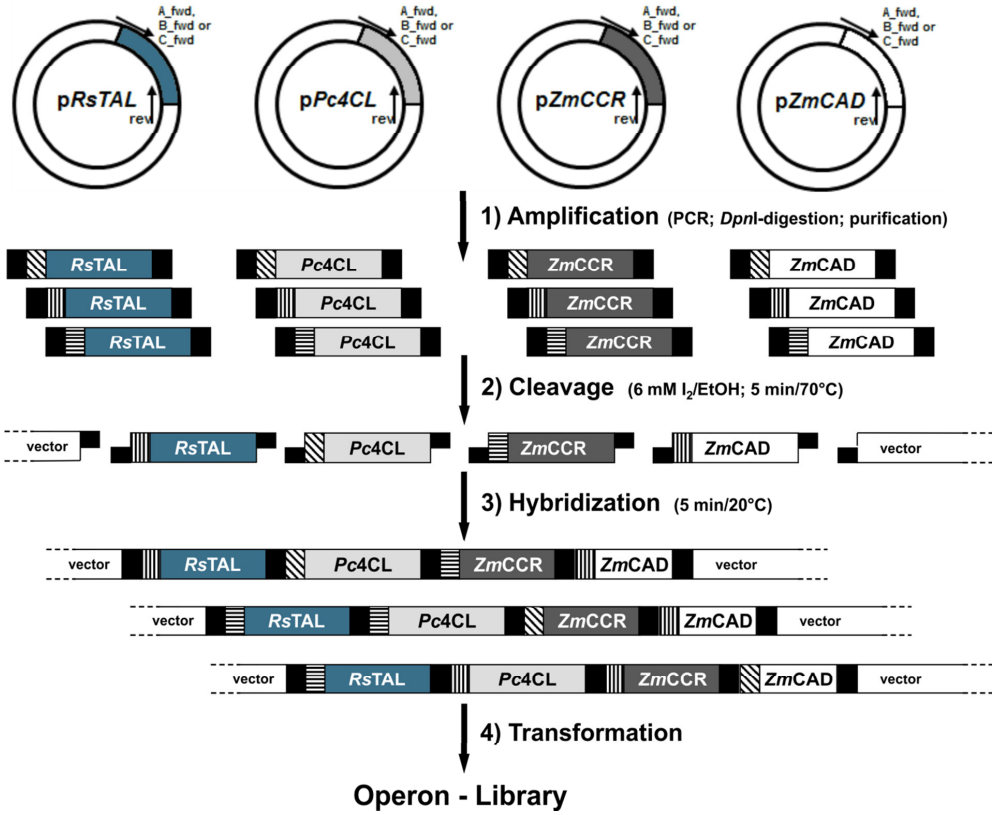


Figure 3. Combinatorial assembly of synthetic *p*-coumaryl-alcohol operons employing the PLICing principle. Three different variants of each gene, having either a 5, 9 or 13 nt SD-sequence - START codon spacing (indicated by different hachures upstream of the respective open reading frames) were generated by PCR with phosphorothioate-oligonucleotides. Subsequently, the phosphorothioate bonds were cleaved for generating complementary 3´-overhangs. The resulting twelve different gene fragments and the vector fragment were mixed in 81 independent hybridization reactions to yield all possible combinations of the tetracistronic *p*-coumaryl-alcohol operon prior to transformation to *E. coli*.

pathway variants to 81 (3^4) different operon configurations as three different SD sequence – START codon spacings are possible for each of the four genes. Simplicity and robustness of the PLICing procedure allowed the individual construction of all 81 variants from PCRs to the final library in a few days (Table 1).

After verification of the completeness of each operon by colony PCR, cultivation and screening in the 96-well plate format for the production of *p*-coumaryl alcohol of this operon library was performed. HPLC-analysis of the supernatant of all strains revealed that the 81 variants can be divided in three distinct groups according to SD sequence –

Table 1. All 81 synthetic *p*-coumaryl alcohol operon variants differing in the combination of the SD sequence-START codon spacing (5 nt, 9 nt, 13 nt) for the genes of the tetracistronic synthetic operon. All variants of the operon were cloned into the pALXtreme vector backbone (*16*) (general scheme: pALXtreme-1a_$T7_{prom}$>x-*Rstal*-x-*Pc4cl*-x-*Zmccr*-x-*Zmcad*) and subsequently transformed to *E. coli* BL21 (DE3) *lacI*^{Q1} (relevant characteristics of the genotype: F^- *ompT hsd*SB(rB^-, mB^-) *gal dcm* (DE3)).

Variant	SD seq.-START spacing	Variant	SD seq.-START spacing	Variant	SD seq.-START spacing
1	5-5-5-5	28	9-5-5-5	55	13-5-5-5
2	5-5-5-9	29	9-5-5-9	56	13-5-5-9
3	5-5-5-13	30	9-5-5-13	57	13-5-5-13
4	5-5-9-5	31	9-5-9-5	58	13-5-9-5
5	5-5-9-9	32	9-5-9-9	59	13-5-9-9
6	5-5-9-13	33	9-5-9-13	60	13-5-9-13
7	5-5-13-5	34	9-5-13-5	61	13-5-13-5
8	5-5-13-9	35	9-5-13-9	62	13-5-13-9
9	5-5-13-13	36	9-5-13-13	63	13-5-13-13
10	5-9-5-5	37	9-9-5-5	64	13-9-5-5
11	5-9-5-9	38	9-9-5-9	65	13-9-5-9
12	5-9-5-13	39	9-9-5-13	66	13-9-5-13
13	5-9-9-5	40	9-9-9-5	67	13-9-9-5
14	5-9-9-9	41	9-9-9-9	68	13-9-9-9
15	5-9-9-13	42	9-9-9-13	69	13-9-9-13
16	5-9-13-5	43	9-9-13-5	70	13-9-13-5
17	5-9-13-9	44	9-9-13-9	71	13-9-13-9
18	5-9-13-13	45	9-9-13-13	72	13-9-13-13
19	5-13-5-5	46	9-13-5-5	73	13-13-5-5
20	5-13-5-9	47	9-13-5-9	74	13-13-5-9
21	5-13-5-13	48	9-13-5-13	75	13-13-5-13
22	5-13-9-5	49	9-13-9-5	76	13-13-9-5
23	5-13-9-9	50	9-13-9-9	77	13-13-9-9
24	5-13-9-13	51	9-13-9-13	78	13-13-9-13
25	5-13-13-5	52	9-13-13-5	79	13-13-13-5
26	5-13-13-9	53	9-13-13-9	80	13-13-13-9
27	5-13-13-13	54	9-13-13-13	81	13-13-13-13

START codon spacing for the gene encoding the RsTAL (Figure 4). This first enzymatic step of the pathway appears to be crucial for the overall pathway performance. All 27 variants (variants 1 - 27) with a 5 nt spacing upstream of the *Rstal* START codon are characterized by higher product concentrations (up to 55 mg/L p-coumaryl alcohol) compared to the 27 variants with a 9 nt spacing upstream of the *Rstal*-gene (variants 28 – 54; accumulation of up to 24 mg/L *p*-coumaryl alcohol). Only up to 9 mg/L *p*-coumaryl alcohol could be detected in the

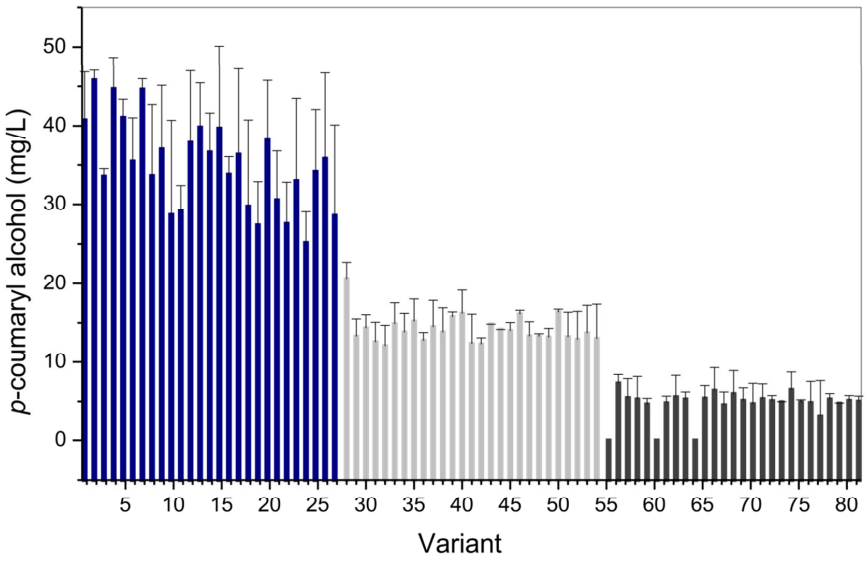


Figure 4. *p*-coumaryl alcohol titers of all 81 constructed variants of the synthetic *p*-coumaryl alcohol operon. The clones are numbered from 1 – 81 in accordance to table 1. The microtiter plate screening was performed in triplicate.

supernatant of clones 55 – 81, all harboring operon variants in which the *Rstal*-gene is characterized by a SD sequence – START codon spacing of 13 nt. These results are in accordance with determined kinetic parameters of various plant derived L-tyrosine ammonia lyases as these enzymes generally display a low catalytic activity (*23-25*). This would explain the need for maximizing the translation efficiency in the genetic environment of the synthetic *p*-coumaryl alcohol operon to compensate for low TAL activity in *E. coli*. Under microtiter plate screening conditions no general trend for the translation efficiency of the other three enzymes was visible. However, during screening, a few variants within the best performing group accumulated more product compared to the 5-5-5-5 variant, indicating that best translation efficiency might not be optimal for highest product titers.

For a deeper insight, shake flask experiments for all 27 clones within this group were performed. The determined *p*-coumaryl alcohol concentrations were put in relation to biomass by simply dividing the determined product titers by the optical density (OD_{600}) of the respective culture at the point of measurement. This was done although only negligible differences in growth could be detected. Unfortunately, again no pattern, identifying optimal translation conditions for the Pc4CL, ZmCCR and ZmCAD could be identified (Figure 5). However, variant 5-5-5-5 was among the poorest producers,

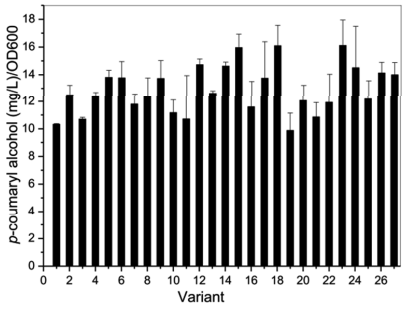


Figure 5. *p*-coumaryl alcohol titers of *E. coli* clones expressing operon variants 1 – 27. Each variant encodes the *Rstal* gene with a SD sequence – START codon spacing of 5 nt. The titers have been corrected for differences in OD_{600} at time of sampling after 12 h. The shake flask experiments were performed in duplicate.

accumulating only up to 10.3 mg/L/OD_{600} *p*-coumaryl alcohol. This supports the notion that a maximum of translation efficiency for all four enzymes of this pathway has a negative effect on product formation. In contrast, for the top five producers 14.6 – 16.1 mg/L/OD_{600} *p*-coumaryl alcohol could be detected in the supernatant after 12 h (variants: no. 12 (5-9-5-13); no. 14 (5-9-9-9); no. 15 (5-9-9-13); no 18 (5-9-13-13); no. 23 (5-13-9-9). A decreased translation efficiency of the other enzymes possibly decreased the metabolic burden for the microbial host with respect to protein synthesis, thus balancing the pathway and probably increasing the availability of the precursor amino acid L-tyrosine for phenylpropanoid synthesis. Extracellular concentrations of *p*-coumaric acid as product of the RsTAL-catalyzed deamination of L-tyrosine and first pathway intermediate were always similar (≤ 3 mg/L), indicating that RsTAL-activity is still limiting the overall product titer despite high expression in *E. coli*. The absolute product titer for these five variants ranged from 47.6 to 52 mg/L *p*-coumaryl alcohol without any further optimization of growth and production conditions. Thus, balancing of the synthetic pathway already more than doubled the overall product titer compared to the previously engineered *E. coli* strain expressing the same genes which accumulated only up to 22 mg/L *p*-coumaryl alcohol without precursor feeding (*20*).

The modularity and simplicity of the PLICing method also offers the opportunity to generate the whole operon library of 81 variants in a single reaction tube by simply combining the 12 amplified and cleaved gene fragments and the equally treated plasmid fragment prior to hybridization and transformation of *E. coli*. Such a library was generated and 300 clones were screened for *p*-coumaryl alcohol production in 96 deep-well plates using three-times oversampling to statistically ensure screening of 95% of all possible variants (*26*). The three individually constructed reference variants (5-5-5-5, 9-9-9-9 and 13-13-13-13) and a strain harboring the empty vector served as controls. Screening and rescreening of the best hits successfully identified the 5-9-9-9 variant, which was also among the best

variants in the manually generated library. However, some good variants were missing raising the question whether the library was complete. DNA sequencing of randomly selected clones of the library surprisingly revealed that not all *E. coli* clones harbored plasmids with a complete synthetic *p*-coumaryl alcohol operon. This was surprising as the five sequences used as points of hybridization during PLICing were all unique and assembly of all five fragements is necessary for circularization. Especially the *Zmcad*-gene was frequently missing. Incomplete pathway assembly was also found during previous studies using the CLIVA method which takes also advantage of the PLICing principle (*27*). Here the number of fully assembled operon variants dropped dramatically when more genes were supposed to be assembled.

Nonetheless, this method for the combinatorial assembly of synthetic pathways described in this manuscript allows the rapid construction of a large number of operon variants. The method is sequence-independent and thus does not require any specific recognition or target sequences for enzymatic activities since all hybridization sites can be arbitrarily selected. In fact, after PCR-amplification, no enzymes such as endonucleases or ligases, which are frequently used in other methods, are needed. The modularity, simplicity and robustness of this method would be perfectly suited for an automation of cloning in the microtiter plate format. During the combinatorial optimization of the *p*-coumaryl alcohol operon, only the SD sequence – START codon spacing was modulated. However, any other regulatory elements of gene expression could also be varied during pathway balancing. Furthermore, five unique hybridization sequences, limiting the overall complexity of the generated library to only 81 variants, were used in this study. Decreasing the number of these sequences (and keeping the number of hybridization sites constant), possibly even using only a single sequence, would allow randomization of the gene order and randomization of the number of genes in the operon library. The latter effect would offer the opportunity to vary the “dosis” of individual genes in the pathway. This would also represent a possible solution to overcome the low TAL activity within the *p*-coumaryl alcohol pathway, which currently limits the product titers in *E. coli*. However, one has to bear in mind that the transformation efficiency drastically decreases with an increasing plasmid size *(20)*. Currently, downside of this method is the observed low cloning efficiency when the five DNA-fragments were assembled in a single reaction tube. Again, the modularity of PLICing could be a solution. When a performed prescreening of an operon library identifies a certain configuration as most beneficial, this parameter can be fixed by simply using other PCR-primer combinations and amplification of different DNA-fragments

prior to construction of another library. In case of the *p*-coumaryl alcohol operon, the 5 nt-*Rstal* fragment, which was identified as being essential for high product titers, can be fused to the vector fragment, thereby reducing the overall number of fragments to be randomly assembled during the construction of the next library of synthetic operons.

Methods

Strains and plasmids. *E. coli* BL21-Gold (DE3) *lacI*^{Q1} and the plasmid pALXtreme-1a were used for cloning and expression purposes (*16*). The cells were routinely cultivated on a rotary shaker (170 rpm) at 37°C in Luria–Bertani (LB) medium or on LB plates (LB medium with 1.5% [wt/vol] agar) (*28*). If appropriate, kanamycin (50 µg/mL) was added. Growth was determined by measuring the optical density at 600 nm (OD_{600}).

Genes and primers. Genes encoding the tyrosine ammonia lyase from *Rhodobacter sphaeroides* (RsTAL), the 4-coumarate: CoA ligase from *Petroselinum crispum* (Pc4CL), the cinnamoyl-CoA reductase from *Zea mays* (ZmCCR) and the cinnamyl–alcohol dehydrogenase, also from *Zea mays* (ZmCAD) were chemically synthesized by GeneArt Gene Synthesis services from Life Technologies, a Thermo Fisher Scientific company (Waltham, MA, USA). Genes were codon optimized for expression in *E. coli* employing the proprietary GeneOptimizer software. Oligonucleotide primers were purchased from Eurofins Genomics (Ebersberg, Germany). Primers used in this study are summarized in table S1 (Supplemental material).

PCR amplification of target genes and vector for PLICing. All genes and the pALXtreme-1a vector backbone were amplified by PCR from plasmid template DNA. PCRs were performed in a final volume of 50 µL in a T3000 thermocycler from Biometra (Göttingen, Germany) using thin walled PCR tubes (Bio-Budget, Krefeld, Germany). Typically, PCR reactions were composed of 1x KOD Hot Start DNA polymerase buffer (Merck Millipore, Billerica, MA, USA), 1.5 mM $MgSO_4$, 0.2 mM dNTP mix, 0.3 µM phosphorothioate forward and reverse oligonucleotides (PTOs) (Table S1), 20 ng template DNA, and 1U KOD Hot Start DNA polymerase. PCR-cycling started with an initial denaturation step at 96°C for 2 min. Subsequently, three cycles (96°C, 20 s; 50°C, 30 s; and 72°C, 120 s), followed by 28 cycles (96°C, 20 s; 55°C, 30 s; and 72°C, 120 s) and one fill-up cycle (72°C, 5 min) were performed. All PCR reactions were subjected to *Dpn*I digestion (10U, 16 h, 37°C) prior to purification of the PCR products using a PCR purification kit (Macherey–Nagel, Düren, Germany) according to manufacturer's instructions. Subsequently, PCR products were quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). For increased PLICing efficiency, a second

PCR reaction with the same composition was performed using the amplified DNA fragments from the first PCR reaction as template. For this purpose, 30 PCR cycles (96°C, 20 s; 55°C, 20 s; and 72°C, 60 s) and one fill-up cycle (72°C, 5 min) after the initial denaturation step (96°C, 2 min) were performed. Again, the PCR products were purified using a PCR purification kit (Macherey–Nagel, Düren, Germany). Agarose–TAE gel electrophoresis was performed according to standard protocols to confirm the presence and correct size of the amplified genes and the pALXtreme-1a vector backbone (*28*).

Construction of the *p*-coumaryl alcohol operon variants using PLICing. The iodine cleavage reactions for PLICing were prepared by mixing 4 µL of the PCR amplified genes (concentrations varied from 0.26 - 0.36 pmol/µL) or 1 µL of the PCR amplified vector backbone DNA, which was adjusted to 4 µL with Milli-Q water (0.06 pmol/µL), with 1 µL cleavage buffer (0.5 M Tris–HCl, pH 9.0), 0.6 µL iodine stock solution (100 mM iodine in 99% ethanol), and 0.4 µL Milli-Q water to a final volume of 6 µl in thin-walled PCR tubes. DNA cleavage was performed by incubating the reaction mixtures at 70°C for 10 minutes (Biometra T3000 thermocycler). The resulting DNA fragments were used directly for hybridization of the DNA fragments without any further purification. DNA hybridization was achieved by sequentially combining the pALXtreme-1a vector backbone, *Rstal*, *Pc4cl*, *Zmccr* and *Zmcad* iodine cleavage mixtures with 5 minutes incubation intervals at room temperature after the addition of each DNA fragment. After a final incubation step at room temperature for 5 minutes, 4 µL of this hybridization mix was used to transform 100 µl chemically competent *E. coli* BL21-Gold (DE3) $lacI^{Q1}$ cells by the RbCl method (*29*).

In order to verify that all four genes of the synthetic *p*-coumaryl alcohol pathway were successfully incorporated into the vector backbone, recombinant clones were analyzed by colony PCR. A master mix with 1x Dream Taq green master mix (Thermo Scientific, Waltham, MA, USA) and 0.3 µM forward and reverse primer (Table S1) was prepared and 20 µl aliquots of this master mix were dispensed into thin walled PCR tubes. Colonies were directly transferred from LB agar plates into the PCR tubes with sterile toothpicks. Cycling started with an initial denaturation at 95°C for 3 min, followed by 30 cycles (95°C, 30 s; 60°C, 30 s; and 72°C, 4 min) and one fill-up cycle (72°C, 5 min). The PCR reactions were analyzed on agarose–TAE gels. The *p*-coumaryl alcohol operon of randomly selected clones was sequenced to confirm the correct assembly of the synthetic operons.

Microtiter plate screening and shake flask cultivations for the microbial production of *p*-coumaryl alcohol. Clones of the generated PLICing-libraries were picked from the LB plates into 96 well microtiter plates (BRAND,

Wertheim, Germany) with 100 µL LB medium per well. These precultures were incubated overnight for 16 hours at 37°C in a microtron shaker from Infors (900 rpm and 75% humidity, [Bottmingen, Switzerland]). 96-well deepwell plates (Eppendorf, Hamburg, Germany) with 600 µL LB medium and 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) per well were inoculated with 6 µl from each preculture. The deepwell plates were incubated at 25°C for 24 hours in the microtron shaker (900 rpm, 75% humidity). The cell-free culture medium was harvested by centrifugation (4°C, 3,500 g, 30 min,) for the determination of product and intermediates. For *p*-coumaryl alcohol production at shake flask scale, 50 mL LB-medium in 500 mL baffled shake flasks was inoculated with 500 µL of an over-night culture and incubated at 37°C and 120 rpm shaking until the OD_{600} reached 0.2. Subsequently, the cultivation temperature was decreased to 25°C and heterologous gene expression was induced with 1 mM IPTG when the OD_{600} reached 0.6. The cultivation continued at 25°C for an additional 24 hours. Samples were taken at regular time intervals for product analysis.

Determination *p*-coumaric acid and *p*-coumaryl alcohol by uHPLC. *p*-coumaryl alcohol and *p*-coumaric acid concentrations in the cell-free culture supernatant were determined by ultra-high-performance LC (uHPLC). For this purpose, 2 µL of the supernatant was injected into an Agilent 1290 infinity LC (Santa Clara, CA, USA) using 0.1% (vol/vol) acetic acid in water (buffer A) and 0.1% (vol/vol) acetic acid in acetonitrile (buffer B) as the mobile phases at a flow rate of 0.4 mL min^{-1}. The LC separation was carried out using a ZORBAX Eclipse AAA (3.5 µm, 4.6 x 75 mm) column with a guard cartridge (4.6 x 12.5 mm) from Agilent at 40°C. For an efficient separation, a gradient was used, where buffer B was increased from 15% to 60% over 10 minutes. Under these conditions, the retention time of *p*-coumaryl alcohol was 4.9 min, wheras *p*-coumaric acid eluted after 5.3 min. The compounds were detected by monitoring the absorption at 275 nm. The *p*-coumaric acid standard was purchased from Sigma-Aldrich (St. Louis, MO, USA), while *p*-coumaryl alcohol was purchased from MicroCombiChem (Wiesbaden, Germany).

Relative protein quantification. Relative protein quantification was performed as previously described (*22*). For the generation of ^{15}N-labelled protein of each target enzyme to serve as internal standards during analyses, cells were cultivated in M9 defined medium supplemented with 50 µM $FeSO_4$ and 1 µM $ZnSO_4$. Glucose (4%) was added as carbon source and $^{15}NH_4Cl$ served as sole nitrogen source. For the generation of protein samples, shake flask cultures were harvested 6 hours after induction with IPTG by centrifugation (4°C, 6,000 g for 10 min). The cell pellets were washed with a

0.9 wt% NaCl solution and centrifuged again (4°C, 6,000 g for 10 minutes). The obtained cells were concentrated 50x in 1 mL lysis buffer [50 mM potassium phosphate buffer, pH 8.0, 2 mM ethylenediaminetetraacetic acid (EDTA), 2 mM dithiothreitol (DTT), supplemented with complete protease inhibitor cocktail (1 697 498, Roche Applied Science)] and shortly incubated on ice. Subsequently, cells were disrupted by sonication for four minutes at 4°C using a Branson sonifier 250 (intensity 2; duty cycle 20%; Branson, [Danbury, CT, USA]). Cell free extracts (CFE) were harvested by centrifugation (4°C, 16,000 g for 30 min).

Author Information

Corresponding Author

* E-mail: j.marienhagen@fz-juelich.de

Author Contributions

J.M. conceived the project and P.S.W. and A.D. designed the experiments. P.S.W. R.V. and S.S. conducted the experiments and S.N. performed statistical data analysis. All authors contributed to the discussion and interpretation of the obtained results. P.S.W and J.M. wrote the manuscript, which was edited by all co-authors.

Notes

The authors declare no competing financial interest.

Acknowledgments

We would like to thank Frank Jansen and Fritz Kreuzaler (RWTH Aachen University) for providing the non-codon optimized genes for the RsTAL, Pc4CL, ZmCCR and ZmCAD for initial experiments and Hubert Ruhrig and Mario Fricke from the mechanical workshop of the IBG-1 (Forschungszentrum Jülich) for the excellent support.

This work was supported by the BOOST-fund “PNP-EXPRESS” from the “NRW Strategieprojekt BioSC” and the ‘‘Helmholtz Initiative Synthetic Biology’’ from the Helmholtz Association.

Abbreviations

SD, Shine-Dalgarno (sequence); PLICing, Phosphorothioate-based ligase-independent gene cloning; TAL, tyrosine ammonia lyase; 4CL, 4-coumarate-CoA-ligase; CCR, cinnamoyl-CoA-reductase; CAD, cinnamyl alcohol dehydrogenase; RsTAL, tyrosine ammonia lyase from *Rhodobacter sphaeroides*; Pc4CL, 4-coumarate-CoA-ligase from *Petroselinum crispum*; ZmCCR, cinnamoyl-CoA-reductase from *Zea mays*; ZmCAD, cinnamyl alcohol dehydrogenase from *Zea mays*.

References

1. Carothers, J. M., Goler, J. A., and Keasling, J. D. (2009) Chemical synthesis using synthetic biology. *Curr. Opin. Biotechnol. 20,* 498-503.

2. Peralta-Yahya, P. P., Zhang, F. Z., del Cardayre, S. B., and Keasling, J. D. (2012) Microbial engineering for the production of advanced biofuels. *Nature 488*, 320-328.
3. Keasling, J. D. (2008) Synthetic biology for synthetic chemistry. *ACS Chem. Biol. 3*, 64-76.
4. Marienhagen, J., and Bott, M. (2013) Metabolic engineering of microorganisms for the synthesis of plant natural products. *J. Biotechnol. 163*, 166-178.
5. Cox, R. S., Surette, M. G., and Elowitz, M. B. (2007) Programming gene expression with combinatorial promoters. *Mol. Syst. Biol. 3*.
6. Hubmann, G., Thevelein, J., and Nevoigt, E. (2014) Natural and modified promoters for tailored metabolic engineering of the yeast *Saccharomyces cerevisiae*. In *Yeast Metabolic Engineering* (Mapelli, V., Ed.), pp 17-42, Springer New York.
7. Burgess-Brown, N. A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., and Gileadi, O. (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expression Purif. 59*, 94-102.
8. Cheadle, C., Fan, J., Cho-Chung, Y. S., Werner, T., Ray, J., Do, L., Gorospe, M., and Becker, K. G. (2005) Stability regulation of mRNA and the control of gene expression. In *Therapeutic Oligonucleotides: Transcriptional and Translational Strategies for Silencing Gene Expression* (ChoChung, Y. S., Gerwirtz, A. M., and Stein, C. A., Eds.), pp 196-204, New York Acad Sciences, New York.
9. Michener, J. K., Thodey, K., Liang, J. C., and Smolke, C. D. (2012) Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab. Eng. 14*, 212-222.
10. Arpino, J. A. J., Hancock, E. J., Anderson, J., Barahona, M., Stan, G. B. V., Papachristodoulou, A., and Polizzi, K. (2013) Tuning the dials of synthetic biology. *Microbiology-Sgm 159*, 1236-1253.
11. Valla, S., and Lale, R., (Eds.) (2014) *DNA cloning and assembly methods*, Vol. 1116, Humana Press, New York.
12. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L. (1995) Single-step assambly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene 164*, 49-53.
13. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343-345.
14. Li, M. Z., and Elledge, S. J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods 4*, 251-256.
15. Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T., and Halkier, B. A. (2007) USER fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res. 35(7): e55*.
16. Blanusa, M., Schenk, A., Sadeghi, H., Marienhagen, J., and Schwaneberg, U. (2010) Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Anal. Biochem. 406*, 141-146.
17. Dennig, A., Shivange, A. V., Marienhagen, J., and Schwaneberg, U. (2011) OmniChange: The sequence independent method for simultaneous site-saturation of five codons. *Plos One 6(10):e26222*.
18. Dennig, A., Marienhagen, J., Ruff, A., and Schwaneberg, U. (2014) OmniChange: Simultaneous site saturation of up to five codons. In *Directed Evolution Library Creation* (Gillam, E. M. J., Copp, J. N., and Ackerley, D., Eds.), pp 139-149, Springer New York.
19. Marienhagen, J., Dennig, A., and Schwaneberg, U. (2012)

Phosphorothioate-based DNA recombination: an enzyme-free method for the combinatorial assembly of multiple DNA fragments. *BioTechniques Rapid Dispatches 0*, 1-6.

20. Jansen, F., Gillessen, B., Mueller, F., Commandeur, U., Fischer, R., and Kreuzaler, F. (2014) Metabolic engineering for *p*-coumaryl alcohol production in *Escherichia coli* by introducing an artificial phenylpropanoid pathway. *Biotechnol. Appl. Biochem.*, Epub Feb 27, 2014. DOI 2010.1002/bab.1222.
21. Chen, H. Y., Bjerknes, M., Kumar, R., and Jay, E. (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* messenger-RNAs. *Nucleic Acids Res. 22*, 4953-4957.
22. Voges, R., and Noack, S. (2012) Quantification of proteome dynamics in *Corynebacterium glutamicum* by ^{15}N-labeling and selected reaction monitoring. *J. Proteomics 75*, 2660-2669.
23. Watts, K. T., Mijts, B. N., Lee, P. C., Manning, A. J., and Schmidt-Dannert, C. (2006) Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. *Chem. Biol. 13*, 1317-1326.
24. Kyndt, J. A., Meyer, T. E., Cusanovich, M. A., and Van Beeumen, J. J. (2002) Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein. *FEBS Lett. 512*, 240-244.
25. Xue, Z. X., McCluskey, M., Cantera, K., Sariaslani, F. S., and Huang, L. X. (2007) Identification, characterization and functional expression of a tyrosine ammonia-lyase and its mutants from the photosynthetic bacterium *Rhodobacter sphaeroides. J. Ind. Microbiol. Biotechnol. 34*, 599-604.
26. Patrick, W. M., Firth, A. E., and Blackburn, J. M. (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng. 16*, 451-457.
27. Zou, R. Y., Zhou, K., Stephanopoulos, G., and Too, H. P. (2013) Combinatorial engineering of 1-deoxy-D-xylulose 5-phosphate pathway using Cross-Lapping *In Vitro* Assembly (CLIVA) method. *Plos One 8*.
28. Sambrook, J., and Russell, D. (2001) *Molecular cloning. A laboratory manual*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
29. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol. 166*, 557-580.

4. Discussion

Phenylpropanoids and phenylpropanoid-derived polyphenols are plant natural products with potential therapeutic applications for the treatment of various cancers, inflammatory diseases, cardiovascular diseases, diabetes and Alzheimer´s disease (18, 74-79). However, once a potent PNP is discovered, the ability to produce enough material for clinical applications becomes the main limiting factor since the plant usually accumulates the desired compound only up to low concentrations. In this context, metabolically engineered microorganisms as production platforms for PNPs represent a very promising alternative to expensive chemical synthesis and cumbersome PNP-extraction from plants (16).

The main goal of this thesis was the development of a microbial production strain for the production of the stilbene pinosylvin from its precursor L-phenylalanine. Pinosylvin has potential application in the treatment of certain cancers, cardiovascular inflammatory diseases and adjuvant arthritis (64-67). Introduction of heterologous pathways into microbial production strains can pose a big challenge due to an unbalanced expression of the individual genes leading to the formation of inclusion bodies or the accumulation of potentially toxic intermediates (68). Hence, an additional goal of this thesis was the advancement of the Phosphorothioate ligase independent gene cloning method (PLICing) (61) to assemble synthetic pathways and balance the heterologous gene expression in bacteria. To achieve this, the PLICing technique was used to construct a library of synthetic operons. In this library the distance between the Shine Dalgarno (SD) sequence and the translation initiation codon of the genes associated with the biosynthetic production of the monolignol *p*-coumaryl alcohol from L-tyrosine, was varied.

Since *E. coli* already proved to be a suitable host for the production of phenylpropanoid derived compounds such as flavonoids (80), stilbenes (81) and coumarins (19), this organism was chosen as platform organism during this work. Construction of an efficient microbial production strain for PNPs requires the identification and selection of suitable enzymes, the assembly of stable genetic constructs for balanced gene expression and the adaptation of the host cell metabolism (pre-cursor/co-factor supply, removal of competing or degradation pathways) to the synthetic pathway (or *vice versa*).

4.1 Microbial pinosylvin production with *Escherichia coli*

Development of stilbene producing microorganisms has been mostly limited to strains requiring the supplementation of phenylpropanoid precursors (e.g. *p*-coumaric acid, *trans*-cinnamic acid, ferulic acid, caffeic acid or sinapic acid). This strategy has led to some considerable success with regard to the stilbene resveratrol, where feeding of *p*-coumaric acid resulted in maximum product titers of 2.3 g/L *in E. coli* when supplementing the fatty acid biosynthesis pathway inhibitor cerulenin (10). While a maximum of only 35 mg/L resveratrol was achieved when supplementing L-tyrosine and malonate, the latter compound being converted to malonyl-CoA through expression of a malonate synthetase (MatB) and malonate carrier protein (MatC) from *Rhizobium trifolii* (82). In comparison, reported product titers for pinosylvin are modest with 0.6 mg/L achieved with *Streptomyces venezuelae* (when supplementing 1.2 mM *trans*-cinnamic acid) (83) and 155 mg/L achieved during biotransformation with *E. coli* (when providing 1 mM *trans*-cinnamic acid and augmenting intracellular malonyl-CoA levels) (84).

In this work several engineering strategies were systematically explored for the construction of a pinosylvin producing *E. coli* strain. Suitable enzymes from various microbial or plant sources for the three-step conversion starting from the amino acid L-phenylalanine were identified through literature and database searches e.g. by employing the *BRaunschweig ENzyme Database* (85) (Figure 4.1).

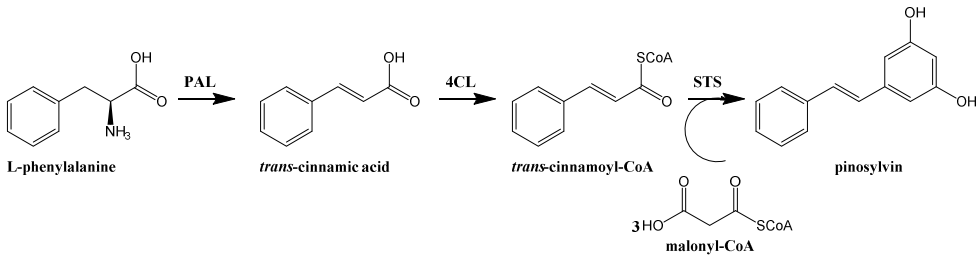


Figure 4.1 Biosynthetic pathway from L-phenylalanine to pinosylvin. *Abbreviations*: PAL, phenylalanine ammonia lyase; 4CL, 4-coumarate-CoA ligase; STS, stilbene synthase.

Two enzyme variants were selected for each of the three enzymatic steps on the basis of their kinetic properties. The phenylalanine ammonia lyase (PAL), 4-Coumarate-CoA ligase (4CL) and stilbene synthase (STS), which catalyze the three subsequent steps from L-phenylalanine to pinosylvin, were initially expressed under the individual control of a T7-

promoter. Soluble expression of the plant STSs in *E. coli* proved to be a difficult task, even though all respective genes were codon optimized for expression in *E. coli* prior to gene synthesis. Heterologous gene expression can be tedious especially when the native expression organism is more distantly related to the heterologous expression host (86, 87). Incomplete folding as a result of absent post-translational modifications and/or the changed microenvironment in the heterologous host or the lack of specific molecular chaperones could hinder heterologous protein expression as it was described for *E. coli* (88). Soluble expression of the STS from *Pinus densiflora* (PdSTS3) was unsuccessful even after attempting multiple expression strategies. During expression of the pinosylvin biosynthetic pathway in *E.* coli, PdSTS3 could be identified in the insoluble protein fraction by SDS-PAGE analysis and the identity of the STS was further confirmed by MALDI-ToF-MS. The N-terminal His_6-tag fusion protein of the *Pinus strobus* STS (PstrSTS2) could be expressed in soluble form and was identified by SDS-PAGE analysis after purification of the fusion protein on a nickel-nitrilotriacetic acid column. Even though the His-tag was initially developed as an affinity tag, it has also been shown to increase the solubility of some small human proteins expressed in *E. coli* (89).

After expression optimization studies, both selected PALs (PAL isozyme 1 from *Petroselinum crispum* [PcPAL1] and the PAL isozyme 2 from *Arabidopsis thaliana* [AtPAL2]) and 4CLs (4CL-like enzyme from *Streptomyces coelicolor* [Sc4CL] and a variant of the 4CL isozyme 2 of *Arabidopsis thaliana* [At4CL2]) were combined with the His_6-tagged fusion of PstrSTS2 (HisPstrSTS2). Each of their respective genes was placed under control of their own T7 promoter to determine the most optimal PAL-4CL-STS combination for pinosylvin production. The combination of PcPAL1 and Sc4CL was the most beneficial for pinosylvin biosynthesis (up to 0.64 mg/L) followed by the AtPAL2 and Sc4CL combination (up to 0.36 mg/L). The positive effect of the N-terminal His_6-tag on PstrSTS2 expression was also reflected by increased pinosylvin titers as the same strain expressing PstrSTS2 without the His_6-tag, accumulated only up to 0.22 mg/L pinosylvin instead of up to 0.64 mg/L. Interestingly, At4CL2 performed better with *A. thaliana* PAL2 (AtPAL2) than with the PAL from *P. crispum* (PcPAL1). The better performance of the AtPAL2/At4CL2 combination could be attributed to the common origin of these enzymes. Moreover these results favor the concept of evaluating the performance of an entire synthetic pathway instead of just assembling a pathway from individually characterized “best parts”. Further improvement of

pinosylvin production could be achieved by organizing the three genes (*Pcpal1*, *Sc*4cl and *HisPstrsts2*) in an operon under control of a single T7 promoter (up to 3.37 mg/L). A previous study also showed improved resveratrol production, when 4CL and STS genes were expressed from an operon-based construct, this possibly led to a more balanced expression according to the authors (81). Replacing the IPTG inducible T7 promoter by the constitutive promoter P_{gap} of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from *E. coli*, however did not result in detectable pinosylvin titers. Low levels of *trans*-cinnamic acid were measured (up to 10 mg/L) for the *E. coli* strain BL21(DE3) pUC18-P*gap-HisPstrsts2-Sc4cl-Pcpal1.* This indicates that all the pathway genes should in principle be expressed as the enzyme product of the last gene in the operon (*Pcpal1*) was detected. Probably expression of the genes is insufficient for detectable pinosylvin production when using the P_{gap}.

During evaluation of the precursor supply, malonyl-CoA availability was identified as the most critical for improving pinosylvin product titers. Addition of the fatty acid biosynthesis pathway inhibitor cerulenin increased the pinosylvin titers 18 to 20-fold (in the absence and presence of L-phenylalanine, respectively). In contrast, L-phenylalanine alone had no positive effect on the product titers. Supply of this amino acid only became limiting during cerulenin-induced increased malonyl-CoA availability. With the aim to provide more of both precursors, it was decided to enhance the intracellular malonyl-CoA levels in the L-phenylalanine *E. coli* production strain 4pF20 (71), which would then serve as host for the production of phenylpropanoid-derived PNPs. Previous metabolic engineering efforts for increasing the intracellular levels of malonyl-CoA in *E. coli*, which included the overexpression of the acetyl-CoA carboxylase (ACC) from *C. glutamicum* and the native acetyl-CoA synthase (ACS), in combination with the deletion of the acetate assimilation pathway (*ackA-pta*) and the ethanol synthesis pathway from acetate (*adhE*) resulted in a 15.7 fold increase of intracellular malonyl-CoA levels (Figure 4.2) (72).

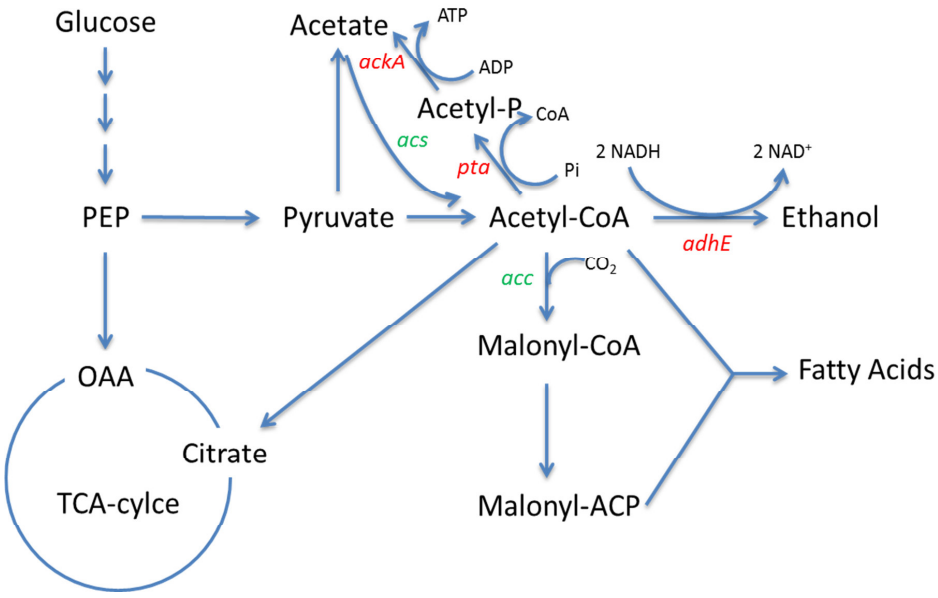


Figure 4.2 Scheme of central metabolism of *E. coli* centered on the biosynthesis of malonyl-CoA. Genes, which were upregulated for the increased synthesis of malonyl-CoA are depicted in green, while genes, which were deleted are depicted in red (72). *Abbreviations*: Acetyl-P, acetyl-phosphate; malonyl-ACP, malonyl-acyl carrier protein; OAA, oxaloacetate; PEP, phosphoenolpyruvate; TCA, tricarboxylic acid.

Unfortunately, the same engineering efforts in *E. coli* 4pF20(DE3), conducted in this thesis did not lead to increased malonyl-CoA or acetyl-CoA levels and simultaneous episomal expression of the pinosylvin pathway and the ACC from *C. glutamicum*, completely abolished the pinosylvin production.

Another strategy to circumvent the addition of cerulenin for increased malonyl-CoA availability was the overexpression of the intrinsic β-ketoacyl-acyl carrier protein (ACP) synthase II (FabF). In a study by Subrahmanyam and Cronan (90), which investigated the inhibitory effect of FabF on phospholipid synthesis, they found that overexpression of FabF increased the ratio of malonyl-CoA to the total cellular CoA pool from 0.5% to 40%. Zha *et al.* (72) investigated various metabolic engineering strategies for improving cellular malonyl-CoA levels in *E. coli* and reported a fourfold increase of malonyl-CoA compared to wild type *E. coli* K-12 upon overexpression of FabF. FabF is believed to sequester most of the malonyl-CoA-ACP transacylase (FabD), blocking the interaction of FabD with the β-ketoacyl-acyl

carrier protein (ACP) synthases I and III (FabB and FabH, respectively). Consequently, FabB and FabH are unable to catalyze essential chain elongation reactions with malonyl-ACP as substrate (Figure 4.3), which subsequently leads to an increase of cellular malonyl-CoA (90). Unfortunately, this strategy also did not lead to increased pinosylvin titers. In a previous study no improvement of the microbial flavonoid production with *E. coli* was observed upon expression of the ACC from *C. glutamicum*, suggesting that at least this strategy is not generally applicable (91).

To investigate whether the availability of the intermediate *trans*-cinnamoyl-CoA could also be limiting due to low 4CL-activity, a comparative analysis of the *trans*-cinnamic acid, *trans*-cinnamoyl CoA and pinosylvin intracellular levels was done. The addition of 200 µM of cerulenin led to an 18-fold increase of pinosylvin, accompanied by a drop of *trans*-cinnamic acid and *trans*-cinnamoyl-CoA level of only 20% and 30% respectively. Thus the PAL-and 4CL-activity in *E.coli* appear to be sufficient to support higher pinosylvin titers, as their intracellular levels are not depleted during pinosylvin biosynthesis. In contrast, expression of the STS was still low even though all STS gene sequences were codon optimized for an expression in *E. coli* prior to gene synthesis. Therefore, the poor soluble expression of HisPstrSTS2 in *E. coli* might be limiting the performance of the pinosylvin pathway.

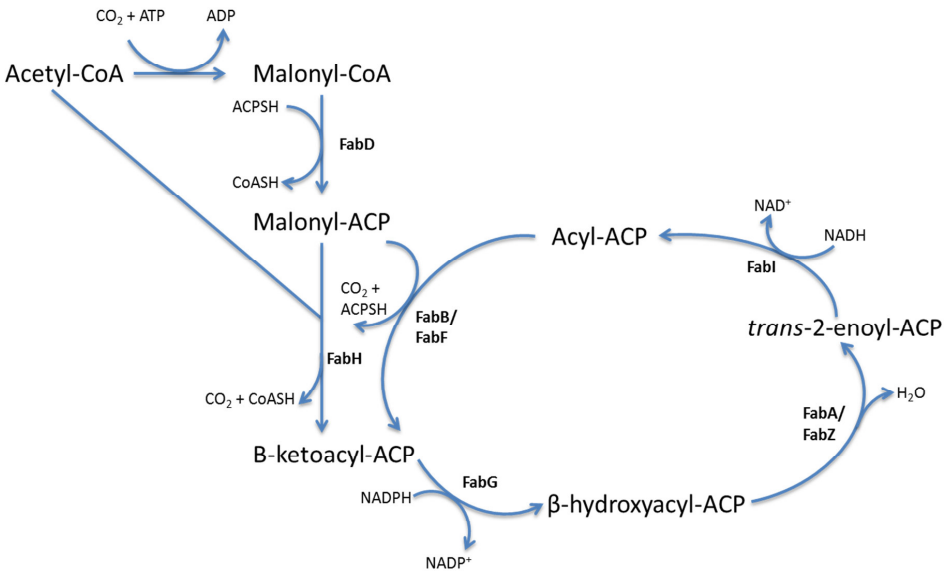


Figure 4.3 The biosynthesis of fatty acids in *E. coli*, with malonyl-CoA and acetyl-CoA as main precursors.

Directed protein evolution was performed to adapt the expression of the pine-derived HisPstrSTS2 to the microbial host system. The codon optimized open reading frame of *HisPstrsts2* was randomly mutated by error prone PCR (epPCR) and the resulting STS-library was subcloned into a plasmid harboring the *Pcpal1* and *Sc4cl* genes. After only one round of mutagenesis by epPCR and MTP-based screening of 450 clones for increased pinosylvin-related fluorescence, two clones with improved product formation were identified. In each clone a single amino acid substitution was identified in their HisPstrSTS2 variants after sequencing of the respective genes. The clone harboring the *HisPstrsts2*-Q361R variant produced 66 mg/L pinosylvin in the presence of 200 µM cerulenin and titers increased to 70 mg/L when L-phenylalanine was also supplemented. The *HisPstrsts2*-T248A variant had the highest pinosylvin titers compared to the best strain *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1*. Pinosylvin titers increased from 59 mg/L measured with *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1* to 70 mg/L with the *HisPstrsts2*-T248A variant in the presence of 200 µM cerulenin, and by an additional 30% to 91 mg/L when L-phenylalanine was also supplemented. However, the combination of both mutations had a detrimental effect on product formation.

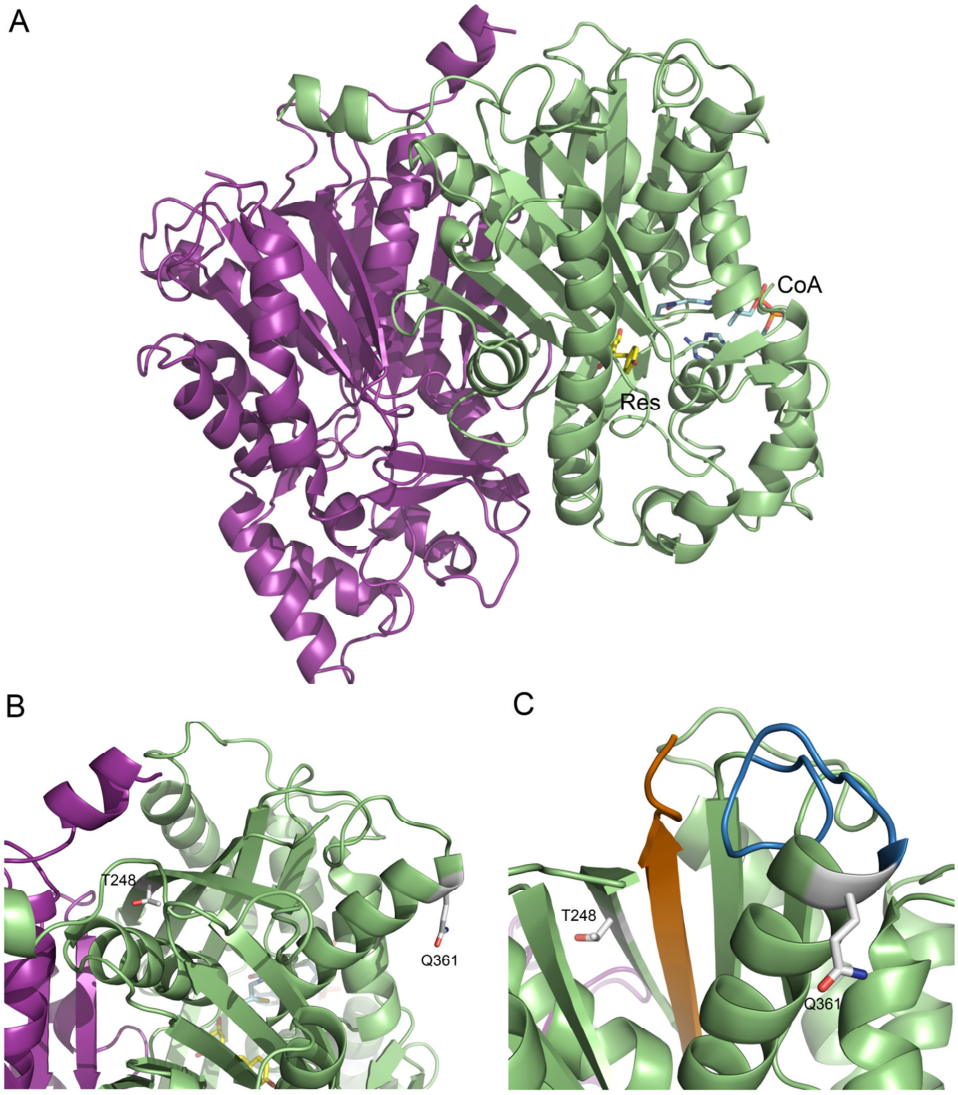


Figure 4.4 (A) Cartoon representation of the PstrSTS2 dimer calculated using the structure of the pinosylvin synthase of *P. sylvestris* as model. The monomers are depicted in green and purple, resveratrol (Res) and coenzyme A (CoA) highlighted in one monomer are shown in the stick mode in yellow and cyan/orange, respectively. (B) T248 is located at the dimer interface, while Q361 is on the protein surface. Both amino acids are shown in the stick mode. (C) T248 and Q361 both interact with the C-terminal β-strand shown in orange. The β-strand, in which T248 is located, is adjacent to this C-terminal β-strand, while the α-helix of Q361 interacts indirectly with the same C-terminal β-strand via a loop shown in blue.

To understand the consequences of the two amino acid substitutions on the structure and activity of the STS, a structure model for a PstrSTS2-dimer was generated using SWISS-MODEL (92) (Figure 4.4A). The crystal structure from the pinosylvin synthase of *P. sylvestris* (PDB code: 1U0U, (93)), which shares an 87% sequence identity to PstrSTS2 on the protein level was used as template for the generation of the PstrSTS2 model. To identify the location of the active site in the PstrSTS2 structural model, a structural alignment with the structures from the closely related chalcone synthase from *Medicago sativa* with coenzyme A bound (PDB code: 1BQ6 (94)) and the stilbene synthase from *Arachis hypogaea* with bound resveratrol (PDB code: 1Z1F (95)) was performed. Both amino acids, T248 and Q361 are not in close proximity to the assumed active site of PstrSTS2 (Figure 4.4B). T248 is located in a β-strand and its side chain is pointing towards the interior of the protein at the dimer interface. Substitution of L-threonine by L-alanine at this position could simply improve the flexibility of the protein and might ultimately lead to the observed improvement of the STS activity in *E. coli.* In contrast, Q361 is positioned at the end of an α-helix on the protein surface and located next to the amino acid sequence N362-G363-C364, which follows the typical *N*-glycosylation sequence N-X-S/T/C of L-asparagine (*N*)-linked protein glycosylation ("X" can be any amino acid except L-proline (96, 97)). However, *E. coli* is not capable of performing L-asparagine (*N*)-linked protein glycosylation. Thus, the substitution of L-glutamine by the more hydrophilic amino acid L-arginine might simply lead to an improved stability or solubility of PstrSTS2 in the absence of any glycosylation in the microbial host. As mentioned previously, combining both amino acid substitutions was detrimental to the pinosylvin production, suggesting that this enzyme variant is inactive. The β-strand, in which T248 is located, is adjacent to the C-terminal β-strand of PstrSTS2, while the α-helix of Q361 interacts indirectly with the same C-terminal β-strand via a loop (Figure 4.4C). Possibly, the combination of both mutations simply destabilizes the structure of PstrSTS2 resulting in an inactive enzyme variant.

4.2 Suggestions to further improve the microbial synthesis of pinosylvin with *E. coli*

In order to achieve significant pinosylvin titers during microbial synthesis with *E. coli*, the addition of cerulenin is still required. With the aim to prevent the addition of this inhibitor multiple genetic alterations of the central carbon metabolism can be carried out to redirect the carbon flux from the glycolysis and citric acid cycle towards malonyl-CoA synthesis. For example naringenin production could be increased by a factor of 5.6, after overexpression of

genes for the pyruvate dehydrogenase multi-enzyme complex (*aceEF* and *lpdA*), the phosphoglycerate kinase (*pgk*), the acetyl-CoA carboxylase (*accABCD*) and deletion of the genes encoding for the fumarase (*fumC*) and succinyl-CoA synthetase (*sucC*). These modifications were predicted by the OptForce model (a computational procedure, which predicts metabolic interventions that lead to the overproduction of a desired biochemical in microbial strains (45)) and led to a 4-fold increase of intracellular malonyl-CoA compared to the wild type strain (98). Combining the genetic alterations in a L-phenylalanine producer strain would directly improve the precursor supply for the STS and PAL enzymes, respectively.

4.3 Combinatorial optimization of synthetic operons for the microbial production of *p*-coumaryl alcohol

In this study, the PLICing method was further developed for the rapid assembly of synthetic pathways. In addition, this method allows for balancing the expression of all pathway genes on the level of translation to maximize product titers by avoiding the accumulation of potentially toxic intermediates and the formation of inclusion bodies. The model pathway for the development of this method was a four-step pathway to convert the amino acid L-tyrosine to the monolignol *p*-coumaryl alcohol (Figure 4.5). The heterologous production of *p*-coumaryl alcohol was recently established in *Escherichia coli*, but without feeding of the precursor *trans-coumaric acid* a final product titer of only 22 mg/L could be achieved (99). Formation of inclusion bodies by the heterologously expressed genes was indicated as the reason for the low product titers.

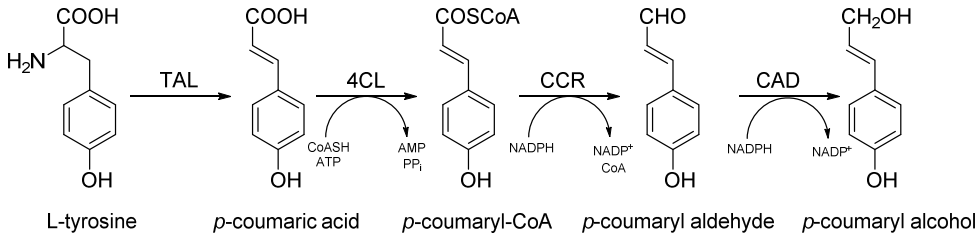


Figure 4.5. Biosynthetic pathway for the production of *p*-coumaryl alcohol from L-tyrosine. *Abbreviations*: TAL, tyrosine ammonia lyase; 4CL, 4-coumarate-CoA ligase; CCR, cinnamoyl-CoA-reductase; CAD, cinnamyl alcohol dehydrogenase.

In the initial experiments, three plasmid-based variants including the four codon optimized genes of the tyrosine ammonia lyase from *Rhodobacter sphaeroides* (*Rstal*), the 4-

coumarate-CoA-ligase from *Petroselinum crispum* (*Pc4cl*), the cinnamoyl-CoA-reductase from *Zea mays* (*Zmccr*) and the cinnamyl alcohol dehydrogenase from *Zea mays* (*Zmcad*) were constructed. In all variants the transcription was controlled by a single IPTG-inducible T7 promoter and the translation of each gene was modulated by an individual, but identical Shine-Dalgarno (SD) sequence (5´-TAAGGAGGT-3´). The three plasmid variants differed only in the spacing between the SD-sequence and the start codon of each gene. The length of this short spacing is known to have significant influence on the translation initiation efficiency (100). While a spacing of 5 nucleotides (nt) confers the optimal translational efficiency, deviation from this optimal length (either shorter or longer than 5 nt) results in stepwise decreased translation efficiency. For the three pathway variants the spacing was set to either 5 nt (5-5-5-5 variant), 9 nt (9-9-9-9 variant) or 13 nt (13-13-13-13 variant) for each gene. The variants were constructed as reference strains with a high, medium and low expression of the four synthetic *p*-coumaryl-alcohol pathway genes, respectively. The relative cytoplasmatic protein concentrations in cell free extract (CFE) determined by targeted proteomics applying isotope dilution mass spectrometry coupled to high performance liquid chromatography (IDMS-LC-MS/MS) verified that the variation of the spacing between the SD sequence and the start codon could indeed control the translation efficiency for each heterologously expressed gene (101). As expected, the highest cytoplasmic protein concentrations were obtained for the 5-5-5-5 variant. With an increasing spacer length between the SD-sequence and start codon the relative cytoplasmic amount of all four proteins decreased stepwise, however, to a different degree for each protein. The level of protein translation was also reflected in the *p*-coumaryl alcohol production of the three reference strains. The 5-5-5-5 variant accumulated up to 40 mg/L *p*-coumaryl alcohol, whereas the 9-9-9-9 variant only accumulated up to 12 mg/l in the supernatant. For the 13-13-13-13 variant the lowest *p*-coumaryl alcohol titer of only 5 mg/L was measured.

After confirming the influence of the SD sequence - START codon spacing on both gene expression and concomitant product formation, a *p*-coumaryl alcohol operon library with variation of the translation efficiency of each gene in the four step pathway was constructed using the PLICing method. For this purpose, three versions of each gene, having either a 5, 9 or 13 nt SD sequence - START codon spacing were manually combined to yield 81 different pathway variants in which the order of the genes was kept constant (*Rstal-Pc4cl-Zmcad-Zmccr*). HPLC-analysis of the supernatant of all 81 strains revealed that the first enzymatic

step of the pathway was the most crucial for the overall pathway performance. Amongst the 27 strains with the highest translation efficiency for RsTAL (5 nt spacing), a titer of 33 to 55 mg/L *p*-coumaryl alcohol could be measured, while the 27 variants with the 9 nt spacing for the *Rstal* gene accumulated up to 24 mg/L p-coumaryl alcohol. For the 27 operon variants, encoding the *Rstal* gene with 13 nt spacing, only up to 9 mg/L *p*-coumaryl alcohol could be detected by HPLC in the supernatant. These results agree with the determined kinetic parameters of various plant derived tyrosine ammonia lyases as these enzymes generally display a low catalytic activity, which would explain the need for a maximization of *Rstal* gene expression in *E. coli* (102-104). No significant trend for the translation efficiency of the three remaining proteins (Pc4CL, ZmCCR and ZmCAD) could be discerned after microtiter plate or shake flask screening experiments. Among the 27 variants with a 5 nt SD sequence-START codon spacing for the *Rstal* gene, the 5-5-5-5 variant was one of the poorest producers accumulating up to 10.3 mg/L/OD600 *p*-coumaryl alcohol, supporting the notion that a maximum expression of all pathway genes has a negative effect on productivity. In comparison, the top five producers accumulated 14.6 – 16.1 mg/L/OD600 *p*-coumaryl-alcohol in the supernatant after 12 h (variants: no. 12 (5-9-5-13); no. 14 (5-9-9-9); no. 15 (5-9-9-13); no 18 (5-9-13-13); no. 23 (5-13-9-9). A decreased translation efficiency of the remaining three enzymes presumably decreased the metabolic burden for *E. coli* with respect to protein synthesis, which in turn balances the pathway and probably increases the availability of L-tyrosine for *p*-coumaryl alcohol synthesis. The levels of *p*-coumaric acid, the product of the RsTAL-catalyzed deamination, measured in the supernatant were always similar and considerably low (≤ 3 mg/L), indicating that the RsTAL-activity in *E. coli* was still limiting the overall product titer.

The absolute product titers for the top five variants measured during shake flask experiments ranged from 47.6 to 52 mg/L without any further optimization of growth and production conditions. Compared to the previously engineered *E. coli* strain expressing the same genes, balancing of the synthetic pathway has already led to a doubling of the *p*-coumaryl alcohol titers (99).

Simultaneous construction of the 81 variants by combining all gene fragments with the vector backbone in a single reaction tube prior to hybridization and transformation to *E. coli*, resulted in the generation of an incomplete library. Even though the 5-9-9-9 variant was

found among the best clones during library screening, DNA sequencing of randomly selected clones revealed that not all *E. coli* clones harbored a complete synthetic *p*-coumaryl alcohol operon. Particularly the *Zmcad* gene was frequently missing. Incomplete pathway assembly was also observed during previous studies employing another phosphorothioate based cloning method (CLIVA), especially when the number of genes to be assembled increased (62).

The PLICing method has proven to be a modular, simple and efficient cloning method, which could be used to set-up an automated cloning procedure in microtiter plate format. However, one has to bear in mind that the assembly efficiency decreases with increasing fragment numbers. To keep the number of fragments manageable, the modularity of the PLICing method can be used. When a certain configuration has been identified as most beneficial for the desired goal during an initial screening round as it was the case for the 5 nt SD sequence – START codon spacing of *Rstal*, this parameter can be fixed during later screening rounds. In case of the *p*-coumaryl alcohol operon, the *Rstal* gene fragment can be incorporated in the vector backbone, thereby reducing the number of fragments to be assembled during the construction of the next generation of synthetic operons.

4.4 Suggestions to balance the heterologous gene expression in microorganisms

During the combinatorial optimization of the *p*-coumaryl alcohol operon, only one parameter namely the SD sequence – START codon spacing was varied. However, other regulatory elements of gene transcription and mRNA translation such as promoters (105) or signal sequences such as metabolite-responsive riboswitches (106) could be varied during pathway balancing. Randomization of the gene order and/ or the number of copies of the individual genes could also represent opportunities for pathway balancing. For example increasing the gene dosage could be a solution to overcome low enzyme activity.

4.5 Conclusions

By combining various metabolic engineering approaches and *in vivo* protein evolution of the heterologously expressed STS from *Pinus strobus*, microbial pinosylvin production from glucose could be successfully established in *E. coli*. However, addition of the fatty acid synthase inhibitor cerulenin is still required for sufficient product titers. Maximum product titers achieved in the presence of 200 µM cerulenin was 70 mg/L, while 91 mg/L could be reached with additional supplementation of 3 mM L-phenylalanine to the medium.

The PLICing method proved to be a fast and efficient technique for the combinatorial assembly of multiple pathway configurations. However for rapid simultaneous construction of operon libraries in a single reaction, this method proved to be unfeasible due to a decreased assembly accuracy of complete operons. Systematic variation of the spacing between the Shine-Dalgarno sequence and the start codon efficiently modulated the expression of individual pathway genes and helped to balance the synthesis of the *p*-coumaryl alcohol pathway enzymes to achieve high product titers. During the evaluation of 81 pathway variants, tyrosine ammonia lyase activity proved to be the limiting step. Maximization of the translation efficiency of the respective mRNA was decisive for the construction of an *E. coli* strain, which accumulated up to 52 mg/L *p*-coumaryl alcohol in the supernatant.

5. References

1. **Marienhagen J, Bott M.** 2013. Metabolic engineering of microorganisms for the synthesis of plant natural products. J. Biotechnol. **163:**166-178.
2. **Bohlmann J, Keeling CI.** 2008. Terpenoid biomaterials. Plant J. **54:**656-669.
3. **Gershenzon J, Dudareva N.** 2007. The function of terpene natural products in the natural world. Nat. Chem. Biol. **3:**408-414.
4. **Pybus D, Sell C.** 2006. The chemistry of fragrances: From perfumer to consumer, 2nd ed. Royal society of chemistry, Cambridge, UK.
5. **Klayman DL.** 1985. Qinghaosu (artemisinin) - an antimalarial drug from China. Science **228:**1049-1055.
6. **Rose WC.** 1992. Taxol - A review of its preclinical *in vivo* antitumor-activity. Anti-Cancer Drugs **3:**311-321.
7. **Kuzuyama T, Seto H.** 2003. Diversity of the biosynthesis of the isoprene units. Nat. Prod. Rep. **20:**171-183.
8. **Miller B, Oschinski C, Zimmer W.** 2001. First isolation of an isoprene synthase gene from poplar and successful expression of the gene in *Escherichia coli*. Planta **213:**483-487.
9. **Takahashi S, Koyama T.** 2006. Structure and function of *cis*-prenyl chain elongating enzymes. Chem. Rec. **6:**194-205.
10. **van der Heijden R, Jacobs DI, Snoeijer W, Hallared D, Verpoorte R.** 2004. The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. Curr. Med. Chem. **11:**607-628.
11. **Julsing MK, Koulman A, Woerdenbag HJ, Quax WJ, Kayser O.** 2006. Combinatorial biosynthesis of medicinal plant secondary metabolites. Biomol. Eng. **23:**265-279.
12. **Halkier BA, Gershenzon J.** 2006. Biology and biochemistry of glucosinolates, p. 303-333, Annu. Rev. Plant Biol., vol. 57. Annual Reviews, Palo Alto.
13. **Sønderby IE, Geu-Flores F, Halkier BA.** 2010. Biosynthesis of glucosinolates - gene discovery and beyond. Trends Plant Sci. **15:**283-290.
14. **Minto RE, Blacklock BJ.** 2008. Biosynthesis and function of polyacetylenes and allied natural products. Prog. Lipid Res. **47:**233-306.
15. **Dembitsky VM.** 2006. Anticancer activity of natural and synthetic acetylenic lipids. Lipids **41:**883-924.
16. **van Summeren-Wesenhagen PV, Marienhagen J.** 2013. Putting bugs to the blush: Metabolic engineering for phenylpropanoid-derived products in microorganisms. Bioengineered **4:**355-362.
17. **Russo GL, Russo M, Spagnuolo C.** 2014. The pleiotropic flavonoid quercetin: from its metabolism to the inhibition of protein kinases in chronic lymphocytic leukemia. Food Funct. **5:**2393-2401.
18. **Pangeni R, Sahni JK, Ali J, Sharma S, Baboota S.** 2014. Resveratrol: review on therapeutic potential and recent advances in drug delivery. Expert Opin. Drug Delivery **11:**1285-1298.
19. **Lin YH, Sun XX, Yuan QP, Yan YJ.** 2013. Combinatorial biosynthesis of plant-specific coumarins in bacteria. Metab. Eng. **18:**69-77.
20. **Kostova I, Bhatia S, Grigorov P, Balkansky S, Parmar VS, Prasad AK, Saso L.** 2011. Coumarins as antioxidants. Curr. Med. Chem. **18:**3929-3951.
21. **Boerjan W, Ralph J, Baucher M.** 2003. Lignin biosynthesis. Annu. Rev. Plant Biol. **54:**519-546.
22. **Korkina L, Kostyuk V, De Luca C, Pastore S.** 2011. Plant phenylpropanoids as emerging anti-inflammatory agents. Mini-Rev. Med. Chem. **11:**823-835.
23. **Davin LB, Lewis NG.** 2005. Dirigent phenoxy radical coupling: advances and challenges. Curr. Opin. Biotechnol. **16:**398-406.
24. **Saarinen NM, Warri A, Airio M, Smeds A, Makela S.** 2007. Role of dietary lignans in the reduction of breast cancer risk. Mol. Nutr. Food Res. **51:**857-866.

25. **Newman DJ, Cragg GM.** 2012. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J. Nat. Prod. **75:**311-335.
26. **Gil- Chávez GJ, Villa JA, Ayala-Zavala JF, Heredia JB, Sepulveda D, Yahia EM, González-Aguilar GA.** 2013. Technologies for extraction and production of bioactive compounds to be used as nutraceuticals and food ingredients: An overview. Compr. Rev. Food Sci. Food Saf. **12:**5-23.
27. **Cho YJ, Hong JY, Chun HS, Lee SK, Min HY.** 2006. Ultrasonication-assisted extraction of resveratrol from grapes. J. Food Eng. **77:**725-730.
28. **Beňová B, Adam M, Pavlíková P, Fischer J.** 2010. Supercritical fluid extraction of piceid, resveratrol and emodin from Japanese knotweed. J. Supercrit. Fluids **51:**325-330.
29. **Staniek A, Bouwmeester H, Fraser PD, Kayser O, Martens S, Tissier A, van der Krol S, Wessjohann L, Warzecha H.** 2013. Natural products - modifying metabolite pathways in plants. Biotechnol. J. **8:**1159-1171.
30. **Lim EK, Bowles D.** 2012. Plant production systems for bioactive small molecules. Curr. Opin. Biotechnol. **23:**271-277.
31. **Weiler EW.** 1997. Octadecanoid-mediated signal transduction in higher plants. Naturwissenschaften **84:**340-349.
32. **Donnez D, Jeandet P, Clément C, Courot E.** 2009. Bioproduction of resveratrol and stilbene derivatives by plant cells and microorganisms. Trends Biotechnol. **27:**706-713.
33. **Wilson SA, Roberts SC.** 2012. Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. Plant Biotechnol. J. **10:**249-268.
34. **Weathers PJ, Towler MJ, Xu JF.** 2010. Bench to batch: advances in plant cell culture for producing useful products. Appl. Microbiol. Biotechnol. **85:**1339-1351.
35. **Schwab W, Davidovich-Rikanati R, Lewinsohn E.** 2008. Biosynthesis of plant-derived flavor compounds. Plant J. **54:**712-732.
36. **Sowbhagya HB, Chitra VN.** 2010. Enzyme-assisted extraction of flavorings and colorants from plant materials. Crit. Rev. Food Sci. Nutr. **50:**146-161.
37. **Newhouse T, Baran PS, Hoffmann RW.** 2009. The economies of synthesis. Chem. Soc. Rev. **38:**3010-3021.
38. **Demain AL, Adrio JL.** 2008. Contributions of microorganisms to industrial biology. Mol. Biotechnol. **38:**41-55.
39. **Du J, Shao ZY, Zhao HM.** 2011. Engineering microbial factories for synthesis of value-added products. J. Ind. Microbiol. Biotechnol. **38:**873-890.
40. **Pickens LB, Tang Y, Chooi YH.** 2011. Metabolic engineering for the production of natural products, p. 211-236. *In* Prausnitz JM (ed.), Annual Review of Chemical and Biomolecular Engineering, Vol 2, vol. 2.
41. **Chemler JA, Koffas MAG.** 2008. Metabolic engineering for plant natural product biosynthesis in microbes. Curr. Opin. Biotechnol. **19:**597-605.
42. **Galm U, Shen B.** 2006. Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. Expert Opin. Drug Discovery **1:**409-437.
43. **Muntendam R, Melillo E, Ryden A, Kayser O.** 2009. Perspectives and limits of engineering the isoprenoid metabolism in heterologous hosts. Appl. Microbiol. Biotechnol. **84:**1003-1019.
44. **Edwards JS, Covert M, Palsson B.** 2002. Metabolic modelling of microbes: the flux-balance approach. Environ. Microbiol. **4:**133-140.
45. **Ranganathan S, Suthers PF, Maranas CD.** 2010. OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comp. Biol. **6**.
46. **Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, Polichuk DR, Teoh KH, Reed DW, Treynor T, Lenihan J, Fleck M, Bajad S, Dang G, Dengrove D, Diola D, Dorin G, Ellens KW, Fickes S, Galazzo J, Gaucher SP, Geistlinger T, Henry R, Hepp M, Horning T, Iqbal T, Jiang H, Kizer L, Lieu B, Melis D, Moss N, Regentin R, Secrest S, Tsuruta H, Vazquez R, Westblade LF, Xu L, Yu M, Zhang Y, Zhao L, Lievense J,**

Covello PS, Keasling JD, Reiling KK, Renninger NS, Newman JD. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature **496:**528-532.

47. **Sambrook J, Russell D.** 2001. Molecular cloning. A laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
48. **Hill R, Eaton-Rye J.** 2014. Plasmid construction by SLIC or sequence and ligation-independent cloning, p. 25-36. *In* Valla S, Lale R (ed.), DNA Cloning and Assembly Methods, vol. 1116. Humana Press.
49. **Wang TW, Ma XY, Zhu H, Li AT, Du GC, Chen J.** 2012. Available methods for assembling expression cassettes for synthetic biology. Appl. Microbiol. Biotechnol. **93:**1853-1863.
50. **Valla S, Lale R.** 2014. DNA cloning and assembly methods, Methods Mol. Biol., vol. 1116. Humana Press, New York.
51. **Hartley JL, Temple GF, Brasch MA.** 2000. DNA cloning using *in vitro* site-specific recombination. Genome Res. **10:**1788-1795.
52. **Zhang YW, Werling U, Edelmann W.** 2012. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. **40**.
53. **Li MZ, Elledge SJ.** 2007. Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. Nat. Methods **4:**251-256.
54. **Aslanidis C, De jong PJ.** 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. **18:**6069-6074.
55. **Schmid-Burgk JL, Xie Z, Frank S, Virreira Winter S, Mitschka S, Kolanus W, Murray A, Benenson Y.** 2012. Rapid hierarchical assembly of medium-size DNA cassettes. Nucleic Acids Res. **40**.
56. **Hsiao KC.** 1993. Exonuclease III induced ligase-free directional subcloning of PCR products. Nucleic Acids Res. **21:**5528-5529.
57. **Zhou MX, Hatahet Z.** 1995. An improved ligase-free method for directional subcloning of PCR amplified DNA. Nucleic Acids Res. **23:**1089-1090.
58. **Bitinaite J, Rubino M, Varma KH, Schildkraut I, Vaisvila R, Vaiskunaite R.** 2007. USER™ friendly DNA engineering and cloning method by uracil excision. Nucleic Acids Res. **35:**1992-2002.
59. **Klock HE, Koesema EJ, Knuth MW, Lesley SA.** 2008. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins: Struct., Funct., Bioinf. **71:**982-994.
60. **Quan JY, Tian JD.** 2009. Circular polymerase extension cloning of complex gene libraries and pathways. Plos One **4**.
61. **Blanusa M, Schenk A, Sadeghi H, Marienhagen J, Schwaneberg U.** 2010. Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. Anal. Biochem. **406:**141-146.
62. **Zou RY, Zhou K, Stephanopoulos G, Too HP.** 2013. Combinatorial engineering of 1-deoxy-D-xylulose 5-phosphate pathway using Cross-Lapping *In Vitro* Assembly (CLIVA) method. Plos One **8**.
63. **Dennig A, Shivange AV, Marienhagen J, Schwaneberg U.** 2011. OmniChange: The sequence independent method for simultaneous site-saturation of five codons. Plos One **6(10):e26222**.
64. **Park EJ, Chung HJ, Park HJ, Kim GD, Ahn YH, Lee SK.** 2013. Suppression of Src/ERK and GSK-3/beta-catenin signaling by pinosylvin inhibits the growth of human colorectal cancer cells. Food Chem. Toxicol. **55:**424-433.
65. **Park EJ, Park HJ, Chung HJ, Shin Y, Min HY, Hong JY, Kang YJ, Ahn YH, Pyee JH, Lee SK.** 2012. Antimetastatic activity of pinosylvin, a natural stilbenoid, is associated with the suppression of matrix metalloproteinases. J. Nutr. Biochem. **23:**946-952.
66. **Jeong E, Lee HR, Pyee J, Park H.** 2013. Pinosylvin induces cell survival, migration and anti-adhesiveness of endothelial cells via nitric oxide production. Phytother. Res. **27:**610-617.
67. **Jančinová V, Perečko T, Nosáľ R, Harmatha J, Smidrkal J, Drábiková K.** 2012. The natural stilbenoid pinosylvin and activated neutrophils: effects on oxidative burst, protein kinase C, apoptosis and efficiency in adjuvant arthritis. Acta Pharmacol. Sin. **33:**1285-1292.

68. **Ajikumar PK, Xiao WH, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G.** 2010. Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. Science **330:**70-74.
69. **Lee N, Francklyn C, Hamilton EP.** 1987. Arabinose-induced binding of AraC protein to the $ara I_2$ activates the *araBAD* operon promoter. Proc. Natl. Acad. Sci. U. S. A. **84:**8814-8818.
70. **Kapust RB, Waugh DS.** 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. **8:**1668-1674.
71. **Gerigk MR, Maass D, Kreutzer A, Sprenger G, Bongaerts J, Wubbolts M, Takors R.** 2002. Enhanced pilot-scale fed-batch L-phenylalanine production with recombinant *Escherichia coli* by fully integrated reactive extraction. Bioprocess Biosystems Eng. **25:**43-52.
72. **Zha WJ, Rubin-Pitel SB, Shao ZY, Zhao HM.** 2009. Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. Metab. Eng. **11:**192-198.
73. **James ES, Cronan JE.** 2004. Expression of two *Escherichia coli* acetyl-CoA carboxylase subunits is autoregulated. J. Biol. Chem. **279:**2520-2527.
74. **Roupe KA, Remsberg CM, Yáñez JA, Davies NM.** 2006. Pharmacometrics of stilbenes: Seguing towards the clinic. Curr. Clin. Pharmacol. **1:**81-101.
75. **Baptista FI, Henriques AG, Silva AMS, Wiltfang J, Slva O.** 2014. Flavonoids as therapeutic compounds targeting key proteins involved in Alzheimer's disease. ACS Chem. Neurosci. **5:**83-92.
76. **Kale A, Gawande S, Kotwal S.** 2008. Cancer phytotherapeutics: Role for flavonoids at the cellular level. Phytother. Res. **22:**567-577.
77. **Gomes A, Fernandes E, Lima J, Mira L, Corvo ML.** 2008. Molecular mechanisms of anti-inflammatory activity mediated by flavonoids. Curr. Med. Chem. **15:**1586-1605.
78. **Rivière C, Richard T, Quentin L, Krisa S, Mérillon JM, Monti JP.** 2007. Inhibitory activity of stilbenes on Alzheimer's beta-amyloid fibrils *in vitro*. Biorg. Med. Chem. **15:**1160-1167.
79. **Wesolowska O, Wisniewski J, Bielawska-Pohl A, Paprocka M, Duarte N, Ferreira MJU, Dus D, Michalak K.** 2010. Stilbenes as multidrug resistance modulators and apoptosis inducers in human adenocarcinoma cells. Anticancer Res. **30:**4587-4593.
80. **Leonard E, Yan Y, Fowler ZL, Li Z, Lim CG, Lim KH, Koffas MAG.** 2008. Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. Mol. Pharm. **5:**257-265.
81. **Lim CG, Fowler ZL, Hueller T, Schaffer S, Koffas MAG.** 2011. High-yield resveratrol production in engineered *Escherichia coli*. Appl. Environ. Microbiol. **77:**3451-3460.
82. **Wu JJ, Liu PR, Fan YM, Bao H, Du GC, Zhou JW, Chen J.** 2013. Multivariate modular metabolic engineering of *Escherichia coli* to produce resveratrol from L-tyrosine. J. Biotechnol. **167:**404-411.
83. **Park SR, Yoon JA, Paik JH, Park JW, Jung WS, Ban YH, Kim EJ, Yoo YJ, Han AR, Yoon YJ.** 2009. Engineering of plant-specific phenylpropanoids biosynthesis in *Streptomyces venezuelae*. J. Biotechnol. **141:**181-188.
84. **Katsuyama Y, Funa N, Miyahisa I, Horinouchi S.** 2007. Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in *Escherichia coli*. Chem. Biol. **14:**613-621.
85. **Schomburg I, Chang A, Schomburg D.** 2014. Standardization in enzymology - Data integration in the world's enzyme information system BRENDA. Perspectives in Science **1:**15-23.
86. **Oeda K, Sakaki T, Ohkawa H.** 1985. Expression of rat-liver cytochrome P-450MC cDNA in *Saccharomyces cerevisiae*. DNA-a Journal of Molecular & Cellular Biology **4:**203-210.
87. **Schwarz F, Huang W, Li CS, Schulz BL, Lizak C, Palumbo A, Numao S, Neri D, Aebi M, Wang LX.** 2010. A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat. Chem. Biol. **6:**264-266.
88. **Baneyx F.** 1999. Recombinant protein expression in *Escherichia coli*. Curr. Opin. Biotechnol. **10:**411-421.

89. **Hammarström M, Hellgren N, van Den Berg S, Berglund H, Härd T.** 2002. Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. Protein Sci. **11:**313-321.
90. **Subrahmanyam S, Cronan JE.** 1998. Overproduction of a functional fatty acid biosynthetic enzyme blocks fatty acid synthesis in *Escherichia coli*. J. Bacteriol. **180:**4596-4602.
91. **Leonard E, Lim KH, Saw PN, Koffas MAG.** 2007. Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. Appl. Environ. Microbiol. **73:**3877-3886.
92. **Arnold K, Bordoli L, Kopp J, Schwede T.** 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22:**195-201.
93. **Austin MB, Bowman ME, Ferrer JL, Schröder J, Noel JP.** 2004. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. Chem. Biol. **11:**1179-1194.
94. **Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP.** 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat. Struct. Biol. **6:**775-784.
95. **Shomura Y, Torayama I, Suh DY, Xiang T, Kita A, Sankawa U, Miki K.** 2005. Crystal structure of stilbene synthase from Arachis hypogaea. Proteins **60:**803-806.
96. **Matsui T, Takita E, Sato T, Kinjo S, Aizawa M, Sugiura Y, Hamabata T, Sawada K, Kato K.** 2011. N-glycosylation at noncanonical Asn-X-Cys sequences in plant cells. Glycobiology **21:**994-999.
97. **Strasser R.** 2014. Biological significance of complex N-glycans in plants and their impact on plant physiology. Front. Plant Sci. **5**.
98. **Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MAG.** 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab. Eng. **13:**578-587.
99. **Jansen F, Gillessen B, Mueller F, Commandeur U, Fischer R, Kreuzaler F.** 2014. Metabolic engineering for *p*-coumaryl alcohol production in *Escherichia coli* by introducing an artificial phenylpropanoid pathway. Biotechnol. Appl. Biochem.**:**Epub Feb 27, 2014. DOI 2010.1002/bab.1222.
100. **Chen HY, Bjerknes M, Kumar R, Jay E.** 1994. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* messenger-RNAs. Nucleic Acids Res. **22:**4953-4957.
101. **Voges R, Noack S.** 2012. Quantification of proteome dynamics in *Corynebacterium glutamicum* by ^{15}N-labeling and selected reaction monitoring. J. Proteomics **75:**2660-2669.
102. **Watts KT, Mijts BN, Lee PC, Manning AJ, Schmidt-Dannert C.** 2006. Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. Chem. Biol. **13:**1317-1326.
103. **Kyndt JA, Meyer TE, Cusanovich MA, Van Beeumen JJ.** 2002. Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein. FEBS Lett. **512:**240-244.
104. **Xue ZX, McCluskey M, Cantera K, Sariaslani FS, Huang LX.** 2007. Identification, characterization and functional expression of a tyrosine ammonia-lyase and its mutants from the photosynthetic bacterium *Rhodobacter sphaeroides*. J. Ind. Microbiol. Biotechnol. **34:**599-604.
105. **Cox RS, Surette MG, Elowitz MB.** 2007. Programming gene expression with combinatorial promoters. Mol. Syst. Biol. **3**.
106. **Michener JK, Thodey K, Liang JC, Smolke CD.** 2012. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. Metab. Eng. **14:**212-222.

6. Appendix

6.1 Supplementary material ''Metabolic engineering of *Escherichia coli* for the synthesis of the plant polyphenol pinosylvin''

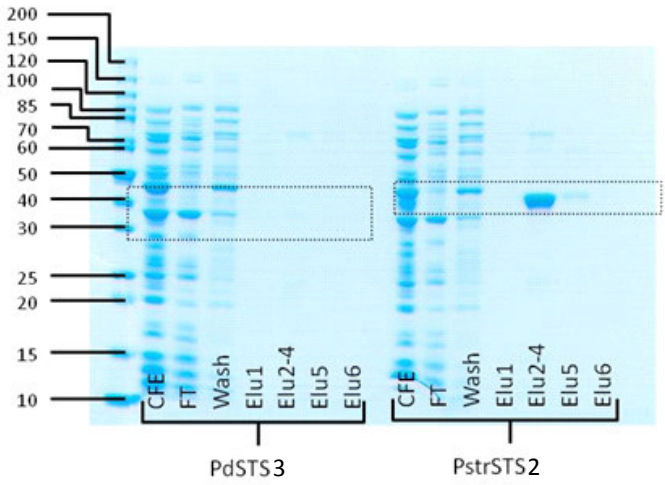


FIG S1 SDS-PAGE analysis of purified HisPdSTS3 and HisPstrSTS2 proteins. Expected molecular weight of HisPdSTS3 is 35.7 kDa and of HisPstrSTS2 is 44.8 kDa. Abbreviations: CFE: Cell free extract before column loading; FT: column flow through; Wash: wash fractions; Elu: Elution fractions.

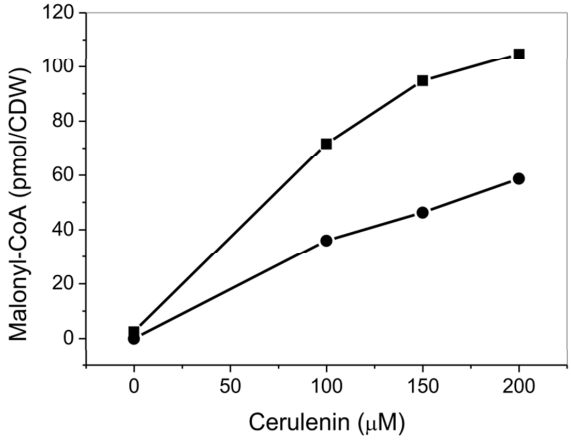


FIG S2 Intracellular malonyl-CoA concentrations in response to different cerulenin concentrations 1.5 hours after addition in *E. coli* BL21(DE3) (■) and the pinosylvin production strain *E. coli* BL21(DE3) pR-*HisPstrsts2-Sc4cl-Pcpal1* (●).

6.2 Supplementary material ''Combinatorial optimization of synthetic operons for the microbial production of *p*-coumaryl alcohol with *Escherichia coli*''

Table S1 Phosphorothioate oligonucleotides used in this study for the construction of the *p*-coumaryl alcohol operon libraries and oligonucleotides for the verification of the correct operon assembly via PCR.

Primer	Sequence[a] (5´-3´)
Phosphorothioate oligonucelotides (PTOs)	
Vec_PTO_fw	AGCaACTAaTAAgCTAAcAAAgCCCGaAAGgAAGCtGAG
Vec_PTO_rv	ACCtCCTtCTTaAAGTTAAACAAAATTATTTCTAGAGGGGAATTG
RsTAL_5_fw	TAAgAAGgAGGtATACCATGCTGGCTATGAGTCCTC
RsTAL_9_fw	TAAgAAGgAGGtGATCATACCATGCTGGCTATGAGTCCTC
RsTAL_13_fw	TAAgAAGgAGGtATCGGATCATACCATGCTGGCTATGAG
RsTAL_uni_rv	CCAcTTAtCCGgATGATTACTAGTGGGTTATTAAACTGGACTCTGTTGC
Pc4CL_5_fw	CCGgATAaGTGgAATAAGGAGGTATACCATGGGAGATTGTGTAGCAC
Pc4CL_9_fw	CCGgATAaGTGgAATAAGGAGGTGATCATACCATGGGAGATTGTGTAGC
Pc4CL_13_fw	CCGgATAaGTGgAATAAGGAGGTATCGGATCATACCATGGGAGATTGTG
Pc4CL_uni_rv	GAAcAGTgGATcATGATTACTAGTGGGTTATTATTTGGGAAGATCACCG
ZmCCR_5_fw	GATcCACtGTTcAATAAGGAGGTATACCATGACCGTCGTCGACGCCG
ZmCCR_9_fw	GATcCACtGTTcAATAAGGAGGTGATCATACCATGACCGTCGTCGACGC
ZmCCR_13_fw	GATcCACtGTTcAATAAGGAGGTATCGGATCATACCATGACCGTCGTC
ZmCCR_uni_rv	CTGaAGCaCTAgATGATTACTAGTGGGTTATTAGGCACGGATGGCG
ZmCAD_5_fw	CTAgTGCtTCAgAATAAGGAGGTATACCATGGGGAGCCTGGCGTCCG
ZmCAD_9_fw	CTAgTGCtTCAgAATAAGGAGGTGATCATACCATGGGGAGCCTGGC
ZmCAD_13_fw	CTAgTGCtTCAgAATAAGGAGGTATCGGATCATACCATGGGGAGCCTG
ZmCAD_uni_rv	AGCtTCCTtTCGgGCTTtGTTaGCTTaTTAgTTGCtGGCCGCATCCGC
Colony PCR primers	
RsTAL_end_fw	GTTCAGGCACTGCGCGAACAGTTTC
ZmCAD_start_rv	ACCAACAACTTTACGTTCGGACGCCAG

[a]Lower case letters mark the locations of phosphorothioate bases.

Acknowledgements

I would like to thank my doctoral advisor Prof. Dr. Michael Bott for giving me the opportunity to work at the IBG-1 on my doctoral thesis and Prof. Dr. Joachim Ernst for agreeing to be my second assessor.

A special thank you goes out to Jan, even though it was and I quote ''still a hard road'', to produce the final thesis at the specified deadline, we made it!!

I would also like to thank my collaboration partners and people who helped obtain much needed data or compounds, plasmids, strains etc.; Raphael Voges, Petra Geilenkirchen, Katharina Neufeld, Alexander Dennig, Melanie Brocker, Sascha Sokolowsky, Stephan Noack, Regina Mahr, Jochem Gätgens, Abigail Koch-Koerfges, Marco Bocola, Alexander Pelzer, Matthias Moch and basically every TA or engineer at IBG-1 (with Karin as my number one go to person of course).

Not to forget the people that made doing a PhD in Jülich fun, were my office mates; Sabrina (also a big thank you for helping me with PhD bureaucracy and thesis formatting issues), Hoffi (my one and only BFFIL), Jenny (I still have fond memories about my Weinachtkalender, maybe this year again?), Katha, Alice, Martin, Lion (I survived without Relentless), Nicolai (ask for a good navigation system this Christmas), Michi and of course the other members of the previous AG Eggeling (yes also including Lothar).

I am grateful to my new Sense up group leaders for giving me a job, chocolate (lots of chocolate, thank you Georg) and a hard deadline for handing in my thesis!

I would also like to thank the people from IT, workshop and administration for helping me with the also important non-lab related stuff!

If you don't see a special shout out to you, don't worry I didn't forget, thank you!

Als laatste en belangrijkste wil ik mijn man Ruben bedanken; voor de ruimte en ondersteuning, zodat ik uiteindelijk een carriere pad kon vinden waar ik mijn ei in kwijt kan.

Band / Volume 90
Ladungstransport durch Graphenschichten und GaAs-Nanodrähte untersucht mit einem Multispitzen-Rastertunnelmikroskop
S. Korte (2014), 96 pp
ISBN: 978-3-89336-990-4

Band / Volume 91
6th Georgian-German School and Workshop in Basic Science
A. Kacharava (Ed.) (2014), CD
ISBN: 978-3-89336-991-1

Band / Volume 92
Ab initio investigations of π-conjugated-molecule-metal interfaces for molecular electronics and spintronics
M. Callsen (2014), viii, 155 pp
ISBN: 978-3-89336-992-8

Band / Volume 93
Ladungstransportmessungen an Si(111) Oberflächen mit einem Multispitzen-Rastertunnelmikroskop
M. Blab (2014), iv, 132, X pp
ISBN: 978-3-89336-997-3

Band / Volume 94
Functional Soft Matter
Lecture Notes of the 46th IFF Spring School 2015
23 February – 06 March, 2015 Jülich, Germany
ed. by J. Dhont, G. Gompper, G. Meier, D. Richter, G. Vliegenthart, R. Zorn
(2015), ca. 600 pp
ISBN: 978-3-89336-999-7

Band / Volume 95
2-Steps in 1-pot: enzyme cascades for the synthesis of chiral vicinal amino alcohols
T. Sehl (2014), XIV, 167 pp
ISBN: 978-3-95806-001-2

Band / Volume 96
Immunohistochemical and electrophysiological characterization of the mouse model for Retinitis Pigmentosa, *rd10*
S. Biswas (2014), XII, 119 pp
ISBN: 978-3-95806-011-1

Band / Volume 97
Single molecule localization microscopy: Imaging of cellular structures and a new three-dimensional localization technique
X. Fan (2014), XII, 92 pp
ISBN: 978-3-95806-014-2

Band / Volume 98
Cryogenic Break-Junction Characterization of Single Organic Molecules
T. Grellmann (2014), VI, 86 pp
ISBN: 978-3-95806-015-9

Band / Volume 99
Interacting Interactions: A Study on the Interplay of Molecule-Molecule and Molecule-Substrate Interactions at Metal-Organic Interfaces
M. Willenbockel (2014), IX, 245 pp
ISBN: 978-3-95806-018-0

Band / Volume 100
Microwire crossbar arrays for chemical, mechanical, and thermal stimulation of cells
P. Rinklin (2015), xii, 184 pp
ISBN: 978-3-95806-022-7

Band / Volume 101
Modification and characterization of potential bioelectronic interfaces
K. Greben (2015), 76 pp
ISBN: 978-3-95806-028-9

Band / Volume 102
Extending the precision and efficiency of the all-electron full-potential linearized augmented plane-wave density-functional theory method
G. Michalicek (2015), 195 pp
ISBN: 978-3-95806-031-9

Band / Volume 103
Metabolic engineering of *Escherichia coli* for the production of plant phenylpropanoid derived compounds
P. V. van Summeren-Wesenhagen (2015), V, 92 pp
ISBN: 978-3-95806-039-5

Weitere ***Schriften des Verlags im Forschungszentrum Jülich*** unter
http://wwwzb1.fz-juelich.de/verlagextern1/index.asp

Schlüsseltechnologien /
Key Technologies
Band / Volume 103
ISBN 978-3-95806-039-5